

ANALISIS GEOSPASIAL KLASTER INOVASI MENDAKAN MAWAR DESA DI KABUPATEN SUKABUMI MENGGUNAKAN K-MEANS DENGAN METODE CRISP-DM

Rani Kusumawati

Program Studi Teknik Informatika Fakultas Sains dan Teknologi
Universitas Muhammadiyah Sukabumi
ranikusumawati2324@ummi.ac.id

ABSTRAK

Inovasi merupakan faktor penting dalam pembangunan daerah, dan Kabupaten Sukabumi memiliki potensi besar dalam inovasi berbasis lokal, salah satunya adalah Mendakan Mawar Desa (MMD), layanan daring administrasi kependudukan yang diselenggarakan oleh Dinas Kependudukan dan Pencatatan Sipil Kabupaten Sukabumi, bekerja sama dengan desa/kelurahan. Penelitian ini bertujuan untuk menganalisis informasi geospasial dan mengidentifikasi pola klaster inovasi MMD di Kabupaten Sukabumi menggunakan algoritma K-Means dalam pendekatan CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Data spasial terkait distribusi geografis dan indikator inovasi MMD dianalisis untuk mengidentifikasi klaster desa berdasarkan tingkat partisipasi dalam program MMD. Hasil analisis menunjukkan klaster desa, yaitu desa yang belum mengajukan kerja sama MMD, tidak aktif, kurang aktif, aktif, dan sangat aktif. Hasil klasterisasi ini kemudian divisualisasikan dalam bentuk informasi spasial berupa peta statis dan peta dinamis, yang memudahkan pemangku kepentingan memantau distribusi partisipasi desa. Temuan ini dapat digunakan sebagai dasar pengambilan keputusan strategis dalam pengembangan dan peningkatan partisipasi desa dalam inovasi MMD berbasis wilayah.

Keyword : *Geospasial, CRISP-DM, Algoritma K-Means.*

1. PENDAHULUAN

Dinas Kependudukan dan Pencatatan Sipil (Disdukcapil) Kabupaten Sukabumi merupakan bagian dari Pemerintah Daerah Kabupaten Sukabumi yang memiliki tanggung jawab di bidang Pendaftaran Penduduk dan Pencatatan Sipil. Tugas utama Disdukcapil adalah mendukung Kepala Daerah, dalam hal ini Bupati Sukabumi, dalam mengelola urusan pemerintahan dan pembangunan yang berkaitan dengan kependudukan dan pencatatan sipil, sesuai dengan peraturan dan undang-undang yang berlaku. Dalam rangka mendukung pengembangan desa yang memiliki peran penting dalam pembangunan daerah, Disdukcapil Kabupaten Sukabumi meluncurkan sebuah inovasi yang dikenal dengan nama Mendakan Mawar Desa (MMD), yang merupakan singkatan dari Mencetak Dokumen Kependudukan Mandiri Warga Desa.

MMD adalah sebuah program yang diselenggarakan oleh Disdukcapil Kabupaten Sukabumi bekerja sama dengan pemerintah desa. Banyaknya desa yang telah bekerja sama memanfaatkan fasilitas layanan administrasi kependudukan daring melalui Inovasi MMD dari mulai diluncurkannya inovasi ini di bulan Desember 2022 sampai dengan bulan Juli 2024 terdapat 122 desa dari total 386 desa dan kelurahan di Kabupaten Sukabumi.

Dalam program ini, desa berperan sebagai operator yang menyelenggarakan layanan administrasi kependudukan secara daring (*online*) bagi warganya. Tujuan utama dari MMD adalah mempermudah dan mempercepat proses

administrasi kependudukan di tingkat desa, sehingga masyarakat dapat lebih mudah mengakses layanan administrasi yang selama ini terbatas oleh jarak atau keterbatasan teknologi. Namun, salah satu tantangan yang dihadapi dalam implementasi MMD adalah bagaimana memetakan dan menganalisis penyebaran serta tingkat keberhasilan inovasi ini di setiap desa. Oleh karena itu analisis geospasial dapat menjadi solusi untuk masalah ini dengan cara memetakan dan mengelompokkan desa-desa berdasarkan implementasi dan hasil dari program MMD.

Untuk melakukan hal tersebut, algoritma K-Means digunakan untuk melakukan klasterisasi atau pengelompokan desa-desa berdasarkan karakteristik dan tingkat keberhasilan penerapan inovasi MMD di tiap desa. Algoritma K-Means merupakan metode yang efektif dalam mengelompokkan data berdasarkan kesamaan atribut, dan sangat relevan untuk analisis data yang berkaitan dengan geospasial dan inovasi.

Selain itu, untuk memastikan analisis dilakukan dengan sistematis dan terstruktur, penelitian ini menggunakan metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*), yang merupakan kerangka kerja yang teruji dalam penerapan teknik data mining untuk mengolah dan menganalisis data. Metode CRISP-DM menyediakan panduan yang jelas melalui langkah-langkah seperti pemahaman masalah, persiapan data, modelisasi, evaluasi, dan penyebaran hasil.

Beberapa penelitian sejenis yang berhubungan dengan pengelompokan data atau topik terkait penggunaan Algoritma K-Means antara lain adalah

“Pemetaan Penduduk Penerima Bantuan Sosial Desa Waru Jaya Menggunakan Algoritma K-Means Clustering” [1]. Penelitian ini menggunakan Algoritma K-Means untuk mengelompokkan penduduk Desa Waru Jaya ke dalam tiga kategori: penduduk yang memenuhi syarat untuk menerima bantuan, penduduk yang kurang memenuhi syarat untuk menerima bantuan, dan penduduk yang tidak memenuhi syarat untuk menerima bantuan. Penelitian lain yang sejenis adalah “Clustering Zonasi Daerah Rawan Bencana Alam di Provinsi Sumatera Barat Menggunakan Algoritma K-Means Clustering dan Library GeoPandas” [2]. Penelitian ini mengelompokkan daerah rawan bencana alam di Provinsi Sumatera Barat menggunakan Algoritma K-Means dan library GeoPandas.

Kajian penelitian terdahulu yang disebutkan di atas memberikan gambaran mengenai bagaimana Algoritma K-Means dapat diterapkan untuk klasterisasi desa berdasarkan berbagai variabel yang mempengaruhi keaktifan dalam pengajuan layanan administrasi kependudukan daring melalui Inovasi Mendakan Mawar Desa (MMD) Mencetak Dokumen Kependudukan Mandiri Warga Desa pada Dinas Kependudukan dan Pencatatan Sipil Kabupaten Sukabumi. Selain itu, penggunaan GeoPandas dalam visualisasi data geospasial membantu memberikan pemahaman yang lebih baik terkait pola-pola geografis yang berhubungan dengan penggunaan layanan publik berbasis teknologi.

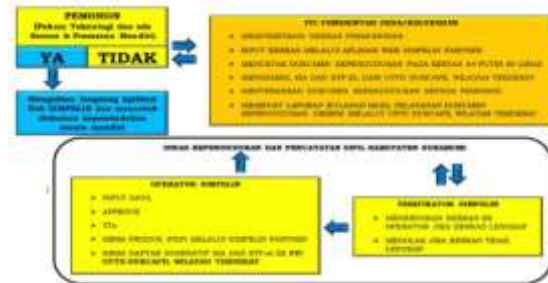
2. TINJAUAN PUSTAKA

2.1. Inovasi MMD

Inovasi Mendakan Mawar Desa (MMD), yang merupakan singkatan dari Mencetak Dokumen Kependudukan Mandiri Warga Desa, adalah sebuah terobosan dalam layanan administrasi kependudukan yang dikembangkan oleh Dinas Kependudukan dan Pencatatan Sipil Kabupaten Sukabumi. Sistem ini memungkinkan masyarakat di desa/kelurahan untuk mengurus administrasi kependudukan secara daring, dengan bantuan petugas desa yang berperan sebagai operator. Dalam implementasinya, petugas desa/kelurahan akan membantu masyarakat dalam proses administrasi kependudukan menggunakan aplikasi web Simpelin Partner (Sistem Informasi Manajemen Pelayanan Online tingkat lanjut), yang dirancang khusus oleh Disdukcapil Kabupaten Sukabumi untuk desa-desa mitra program MMD.

Dalam layanan ini, masyarakat hanya perlu datang ke kios pelayanan administrasi kependudukan yang tersedia di desa/kelurahan dan membawa dokumen-dokumen yang diperlukan untuk pembuatan dokumen kependudukan. Selanjutnya, petugas desa yang ditunjuk sebagai PIC (*Person In Charge*) atau operator MMD oleh Kepala Desa akan memfasilitasi proses pendaftaran secara online, menginput data, serta mengunggah dokumen

yang diperlukan untuk menyelesaikan permohonan administrasi kependudukan. Melalui inovasi ini, masyarakat dapat memperoleh dokumen kependudukan dengan lebih cepat dan efisien, tanpa harus bepergian jauh ke kantor Disdukcapil/UPTD Dukcapil.



Gambar 1. Alur Inovasi MMD

2.2. K-Means Clustering

Algoritma K-Means Clustering adalah metode yang umum digunakan dalam analisis data dan pembelajaran mesin, dengan tujuan untuk mengelompokkan data ke dalam beberapa kluster berdasarkan kesamaan yang terdapat di dalamnya. K-Means adalah metode yang sederhana untuk analisis klasterisasi yang bertujuan untuk mencari cara terbaik dalam membagi entitas ke- n menjadi kelompok yang disebut kluster.

2.3. Langkah-langkah Algoritma

Langkah pertama adalah memilih jumlah kluster yang diinginkan, kemudian secara acak menentukan titik pusat (centroid) awal untuk setiap kluster. Setelah itu, pilih K titik data secara acak sebagai centroid awal dari kluster-kluster tersebut. Setiap titik data x_i akan ditempatkan pada kluster yang memiliki centroid terdekat, yang dihitung menggunakan jarak Euclidean dengan rumus berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Rumus di atas menjelaskan tentang jarak Euclidean antara dua titik, yang dilambangkan dengan $d(x, y)$, dimana x_i dan y_i merupakan komponen ke- i dari titik x dan y dan n adalah jumlah dimensi ruang tempat kedua titik berada. Selanjutnya, $(x_i - y_i)^2$ merupakan kuadrat dari selisih antara komponen-komponen titik x dan y , sementara $\sum_{i=1}^n (x_i - y_i)^2$ adalah penjumlahan dari kuadrat selisih antara komponen-komponen titik x dan y untuk seluruh dimensi. Akhirnya $\sqrt{\cdot}$ menunjukkan akar kuadrat dari jumlah kuadrat selisih komponen-komponen tersebut.

Setelah setiap titik data ditempatkan pada kluster yang sesuai, hitung kembali posisi centroid kluster dengan mengambil rata-rata posisi semua titik dalam kluster tersebut. Ulangi (iterasi) langkah

ini hingga posisi centroid stabil dan tidak berubah secara signifikan, atau hingga jumlah iterasi maksimum tercapai. Ketika konvergensi tercapai (yaitu centroid tidak berubah), algoritma berhenti. Hasil akhirnya adalah pembagian data ke dalam K kluster dengan centroid yang stabil.

Salah satu teknik untuk menentukan jumlah kluster yang optimal dengan membandingkan persentase hasil antara jumlah kluster yang membentuk siku pada suatu titik adalah menggunakan Metode Elbow dengan menggunakan rumus berikut :

$$SSE = \sum_{i=1}^n \sum_{k=1}^K I(c_i = k) \cdot |x_i - \mu_k|^2 \quad (2)$$

Rumus (2) di atas menjelaskan bahwa n adalah jumlah titik data, K adalah jumlah kluster, dan c_i adalah indeks kluster yang ditugaskan pada titik data ke- i . Fungsi indikator $I(c_i = k)$ bernilai 1 jika titik data ke- i berada dalam kluster k , dan 0 jika tidak. x_i adalah titik data ke- i , sementara μ_k adalah pusat (*centroid*) kluster ke- k . Jarak kuadrat antara titik data x_i dan centroid kluster μ_k dilambangkan dengan $|x_i - \mu_k|^2$.

Yang Terakhir, mengevaluasi kualitas hasil klusterisasi data dengan menggunakan metrix *Rand Indeks* untuk mengukur persentase pasangan data points yang dikelompokkan dengan tepat dalam suatu klusterisasi. Nilai *Rand Index* berkisar antara 0 hingga 1, di mana nilai yang lebih tinggi menunjukkan hasil klusterisasi yang lebih baik, menggunakan rumus berikut :

$$RI = \frac{a + b}{\binom{n}{2}} \quad (3)$$

Berikut terdapat beberapa keterangan pada rumus (3) di atas, diantaranya a adalah jumlah pasangan elemen yang ada dalam kluster yang sama pada kedua pengelompokan, sedangkan b adalah jumlah pasangan elemen yang berada pada kluster yang berbeda pada kedua pengelompokan. Selain itu, $\binom{n}{2}$ adalah jumlah pasangan yang mungkin antara dua elemen dalam dataset, dihitung sebagai kombinasi dua elemen dari total n elemen.

2.4. Informasi Spasial

Informasi Geospasial, yang selanjutnya disingkat IG, adalah Data Geospasial (DG) yang telah diolah sehingga dapat digunakan sebagai alat bantu dalam perumusan kebijakan, pengambilan keputusan, dan/atau pelaksanaan kegiatan yang berkaitan dengan ruang kebumih. Adanya informasi geospasial ini akan membantu untuk menentukan keputusan yang tepat, efisien, dan juga komunikatif [3].



Gambar 2. Contoh Persebaran Covid-19

2.5. Python

Python adalah bahasa pemrograman tingkat tinggi karena keunggulannya yang memiliki banyak modul, pustaka, dan framework yang tersedia, sehingga memudahkan penggunaannya karena hanya perlu mengimpor atau menyesuaikan yang sudah ada. Selain itu, Python sangat fleksibel, cocok untuk proyek kecil maupun besar, dapat digunakan secara offline maupun online, serta mendukung pengembangan proyek dengan atau tanpa antarmuka pengguna grafis (GUI).

2.6. Geopandas

Geopandas adalah salah satu library *open-sources* pada Python yang berfungsi untuk memanipulasi dan memvisualisasikan data geospasial. Geopandas menggunakan struktur data inti berupa GeoDataFrame untuk memanipulasi dan memvisualisasikan data [4].

2.7. Folium

Folium adalah pustaka Python yang membantu membuat beberapa jenis peta Leaflet. Karena hasil Folium bersifat interaktif, pustaka ini sangat berguna untuk membangun dasbor. Dengan menggunakan Folium, kita dapat memanipulasi data dalam Python, lalu memvisualisasikannya dalam peta Leaflet.

2.8. Flask

Flask adalah sebuah framework web yang dibangun menggunakan bahasa pemrograman Python. Seperti halnya framework lainnya, Flask memudahkan pengembang dalam membuat aplikasi dengan menyediakan template skrip yang dapat digunakan kapan saja. Flask termasuk dalam kategori micro framework, yang berarti merupakan framework minimalis [5].

2.9. Quantum Geographic Information System (QGIS)

QGIS adalah perangkat lunak SIG sumber terbuka yang memungkinkan pengguna untuk melakukan berbagai tugas yang berkaitan dengan data geografis.

2.10. Visual Studio Code

Visual Studio Code adalah editor source code yang bebas biaya, ringan, dan bersifat open source, yang dikembangkan oleh Microsoft. Editor ini sangat cocok digunakan untuk pengembangan dengan JavaScript, TypeScript, dan Node.js.

Dirancang untuk kesederhanaan, kecepatan, serta kenyamanan bagi pengembang individu maupun tim kecil, Visual Studio Code menjadi pilihan yang populer di kalangan developer untuk berbagai macam tugas pemrograman.

2.11. CRISP-DM

Cross-Industry Standard Process for Data Mining (CRISP-DM), menyediakan sebuah proses standar untuk data mining sebagai strategi pemecahan masalah yang berlaku secara umum dalam bisnis maupun penelitian. Pada CRISP-DM, proses data mining terdapat siklus hidup yang terdiri dari enam tahap yang saling berurutan dan fleksibel. Fase atau tahap pertama adalah Fase Pemahaman Bisnis, kemudian Fase Pemahaman Data, Fase persiapan data, Fase pembuatan model, fase evaluasi, model dan yang terakhir adalah fase penyebaran/development [6].



Gambar 3. Fase Dalam Metode CRISP-DM

3. METODE PENELITIAN

3.1. Data dan Sumber Data

Dalam penelitian ini, dua jenis data utama digunakan sebagai dasar analisis, yaitu data tabular yang berkaitan dengan pengajuan administrasi kependudukan dan data geospasial yang berkaitan dengan lokasi geografis desa. Data administratif yang digunakan berasal dari Sistem Informasi Manajemen Pelayanan Administrasi Kependudukan (SIMPELIN Partner) yang dikelola oleh Dinas Kependudukan dan Pencatatan Sipil Kabupaten Sukabumi. Selain itu, penelitian ini juga memanfaatkan data geospasial yang mencakup geometri desa-desa di Kabupaten Sukabumi, yang diperoleh dari www.indonesia-geospasial.com. Data geospasial ini mencakup koordinat geografis setiap desa yang digunakan untuk analisis spasial dan visualisasi.

3.2. Tahapan Penelitian

Proses analisis data mengikuti tahapan metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*), yang meliputi langkah-langkah berikut:

3.3. Tahap 1 : Pemahaman Bisnis

Pada tahap ini, fokus utamanya adalah memahami esensi dari proses data mining serta kebutuhan yang berasal dari perspektif bisnis. Penelitian ini fokus meningkatkan layanan administrasi kependudukan di desa dengan menggunakan algoritma K-Means Clustering untuk mengelompokkan desa berdasarkan keaktifan layanan daring, membuat peta pemantauan, dan membantu keputusan lebih cepat dengan data yang terstruktur, sekaligus mengatasi masalah layanan tidak merata dan sulitnya mengelola sumber daya serta data peta.

3.4. Tahap 2 : Pemahaman Data

Pada tahap ini, data yang mempunyai keterkaitan dikumpulkan dan dianalisis agar memperoleh wawasan mengenai karakteristik dan kualitasnya. Proses ini juga melibatkan eksplorasi awal untuk mengidentifikasi pola dan mendeteksi anomali, yang berguna untuk menguji hipotesis serta memverifikasi asumsi dengan menggunakan statistik deskriptif dan visualisasi grafis. Pada tahap Pemahaman Data ini, diasumsikan bahwa data telah tersedia dan siap untuk dianalisis lebih mendalam.

3.5. Tahap 3 : Persiapan Data

Tahap ini sering disebut sebagai tahap yang memerlukan banyak pekerjaan. Aktivitas yang dilakukan meliputi pemilihan tabel dan field yang akan ditransformasikan ke dalam database baru yang akan digunakan untuk data mining.

3.6. Tahap 4 : Pemodelan Data

Modeling adalah tahap yang melibatkan pemilihan teknik data mining yang akan digunakan, serta penentuan tools, algoritma, dan parameter yang optimal. Dalam penelitian ini, algoritma yang diterapkan adalah K-Means Clustering, sebuah algoritma *Unsupervised Learning* yang sering digunakan untuk mengelompokkan data berdasarkan kesamaan fitur [7]. Algoritma K-Means Clustering bekerja dengan membagi data ke dalam beberapa klaster yang memiliki kesamaan dalam karakteristik tertentu, seperti halnya keaktifan dalam pengajuan layanan administrasi kependudukan.

3.7. Tahap 5 : Evaluasi Model

Tahap evaluasi merupakan langkah krusial yang memastikan bahwa model yang dibuat mampu menghasilkan output yang akurat dan mencapai tujuan yang telah ditetapkan.

3.8. Tahap 6 : Penyebaran/Deployment

Tahap selanjutnya dalam proses CRISP-DM adalah *deployment*. Pada tahap ini, model yang telah dihasilkan melalui analisis dan pengolahan data sebelumnya akan disebar atau diterapkan untuk digunakan dalam proses selanjutnya. Penyebaran model ini dilakukan agar pihak-pihak terkait, seperti

pengguna akhir, tim pengambil keputusan, atau instansi lain, dapat memanfaatkan hasil model untuk tujuan tertentu, seperti pengambilan keputusan, prediksi, atau optimasi dalam kegiatan mereka. *Deployment* memastikan bahwa model yang telah dibuat dapat diterapkan dalam situasi dunia nyata, memberikan manfaat praktis bagi organisasi atau pihak yang terlibat.

3.9. Algoritma K-Means Clustering

Proses klusterisasi dilakukan dengan algoritma K-means, yang memerlukan penentuan jumlah kluster yang optimal. Jumlah kluster ditentukan berdasarkan analisis nilai *Within-Cluster Sum Of Squares (WCSS)*, yang digunakan untuk meminimalkan variasi dalam kluster dan memaksimalkan variasi antar kluster [8].

4. HASIL DAN PEMBAHASAN

4.1. Pemahaman Bisnis

Penelitian ini bertujuan untuk menganalisis klusterisasi desa dalam program inovasi Mendakan Mawar Desa (MMD) di Kabupaten Sukabumi berdasarkan data administratif pengajuan layanan daring dan data geospasial. Melalui klusterisasi, diharapkan dapat ditemukan pola distribusi inovasi yang dapat digunakan untuk mengukur keberhasilan program MMD. Hasil analisis tersebut akan divisualisasikan dalam bentuk peta dinamis yang diharapkan dapat mempermudah pengambilan keputusan dalam merancang kebijakan berbasis wilayah yang lebih tepat sasaran.

4.2. Mengumpulkan Data

Data yang digunakan untuk klusterisasi mencakup pengajuan layanan administrasi kependudukan daring, yang meliputi total pengajuan yang masuk, pengajuan yang ditolak, dan pengajuan yang diproses. Periode data yang diambil mencakup sejak dimulainya kerjasama dengan desa dengan Disdukcapil hingga *cut-off* di tanggal 25 Juli 2024.

Tabel 1. Sample data pengajuan layanan adminduk daring oleh desa

Desa / Kecamatan	Pengajuan		
	Total	Ditolak	Selesai
Bantargadung/ Bantargadung	85	25	59
Bantargebang/ Bantargadung	404	207	191
Bojonggaling/ Bantargadung	1	1	0
Boyong Sari/ Bantargadung	87	24	60
Buanajaya/ Bantargadung	187	100	84
Limusnunggal/ Bantargadung	0	0	0
Mangunjaya/ Bantargadung	2	2	0
Berkah/ Bojonggenteng	8	3	5

Untuk pengumpulan data spasial tahapan yang pertama dilakukan adalah memisahkan data SHP (*Shapefile*) Kabupaten/Kota pada peta provinsi Jawa Barat hanya wilayah Kabupaten Sukabumi saja, kemudian hasilnya disimpan dalam file GeoJSON.



Gambar 4. Peta Kabupaten Sukabumi

Fitur data spasial yang digunakan dalam penelitian ini meliputi kode_desa, desa, kecamatan, dan geometry.

Setelah seluruh dataset terkumpul, selanjutnya memproses dataset dengan menggunakan bahasa pemrograman python dan editor code visual studio code. Dataset tersebut kemudian diload dan digabung (*merge*).

Gambar 5. Dataset Setelah Digabung

4.3. Mendeskripsikan Data

Langkah memahami dan mendeskripsikan karakteristik dari data yang telah dikumpulkan. Informasi yang akan dihasilkan meliputi: *count* untuk jumlah nilai yang tidak kosong, *mean* untuk nilai rata-rata, *std* untuk simpangan baku.

	pengajuan_total	pengajuan_selesai	pengajuan_ditolak
count	386.000000	386.000000	386.000000
mean	21.940415	13.432642	7.891192
std	74.832760	49.924482	25.030549
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000
50%	0.000000	0.000000	0.000000
75%	0.000000	0.000000	0.000000
max	725.000000	500.000000	207.000000

Gambar 6. Deskriptif Dataset

4.4. Memahami Data

Explore Data adalah untuk menggali lebih dalam dan mengidentifikasi pola atau hubungan yang lebih rumit antar variabel atau untuk menguji

hipotesis awal yang mungkin muncul setelah melihat statistik deskriptif. Eksplorasi ini juga dapat melibatkan visualisasi data untuk lebih memahami data secara intuitif



Gambar 7. Statistik Data Kerjasama MMD

Berikut adalah visualisasi grafik kerjasama inovasi MMD dapat dilihat dalam gambar dibawah



Gambar 8. Grafik Data Kerjasama MMD

4.5. Menghilangkan Data Missing

Jika ditemukan data yang hilang, data tersebut akan dihapus. Berdasarkan pemeriksaan, tidak ditemukan data yang hilang dalam penelitian ini, sehingga proses pra-pemrosesan dapat dilanjutkan ke tahap berikutnya.

4.6. Standarisasi Data

Standarisasi data dalam penelitian ini menggunakan *minmaxscaler*, yang bertujuan untuk mengubah data ke dalam rentang yang lebih konsisten, biasanya antara 0 dan 1, sehingga seluruh data berada dalam skala yang seragam.

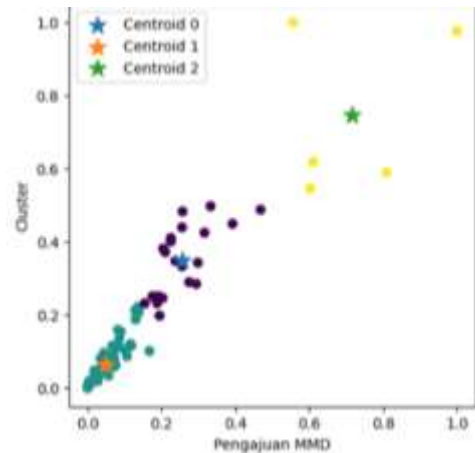
4.7. Pemodelan Data

Algoritma K-Means untuk mengklusterkan desa-desa berdasarkan atribut administratif dan geospasial. Listing Program Python yang digunakan untuk proses ini sebagai berikut:

```
110.
111. # Klusterisasi dengan jumlah kluster awal 3
112. k_means = KMeans(n_clusters=3, init='k-means++',
113.                  random_state=42)
114.
115. # Assign a label for each data point based on their cluster
116. y_kluster = k_means.fit_predict(x_train)
117. y_kluster
118.
119.
120. array([[1, 1, 0, 1, 0, 1, 0, 1, 1, 2, 0, 0, 1, 1, 1, 1, 1, 1, 2, 1, 1, 0,
121.         1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 0, 0,
122.         1, 1, 1, 1, 0, 1, 1, 1, 1, 2, 0, 0, 1, 1, 1, 1, 0, 0, 1, 1,
123.         1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1], dtype=int32)
```

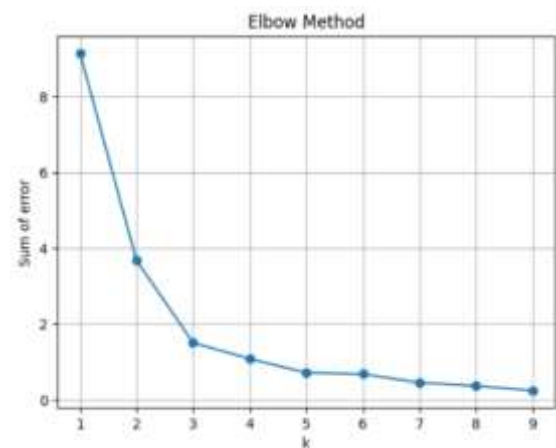
Gambar 9. Kode Program K-Means Clustering

Adapun hasil visualisasi dari algoritma K-Means Clustering menggunakan rumus diatas :



Gambar 10. Visualisasi Hasil Klusterisasi

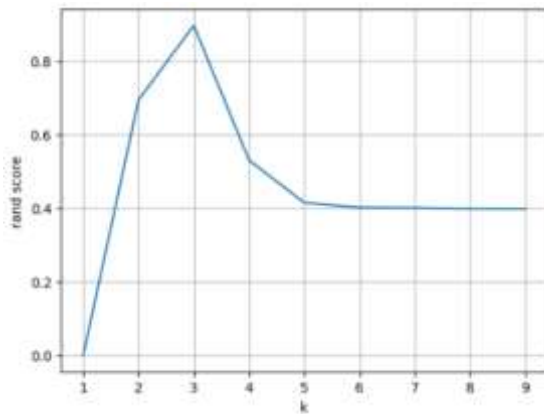
Penentuan jumlah kluster yang optimal dilakukan dengan menggunakan metode elbow untuk memilih jumlah kluster yang menghasilkan pembagian yang paling efisien dan signifikan



Gambar 11. Metode Elbow Untuk Cluster Terbaik

4.8. Evaluasi Model

Evaluasi model menggunakan Indeks Rand, yang memiliki rentang nilai antara 0 dan 1, di mana nilai 1 mengindikasikan kecocokan yang sempurna antara dua partisi, sementara nilai yang mendekati 0 menunjukkan penugasan acak titik data ke dalam kluster [9], berikut hasil visualisasinya:



Gambar 12. Evaluasi Kluster Dengan Rand Index

Berdasarkan hasil klasterisasi menggunakan algoritma K-Means, ditemukan beberapa kluster seluruh desa sebagai berikut: 264 desa tidak mengajukan kerjasama MMD, 40 desa mengajukan kerjasama MMD dengan status tidak aktif, 56 desa mengajukan kerjasama MMD dengan status kurang aktif, 21 desa mengajukan kerjasama MMD dengan status aktif, dan 5 desa mengajukan kerjasama MMD dengan status sangat aktif.

	Status Cluster	Pengajuan Rate-Rata Minimal	Pengajuan Rate-Rata Maksimal	Jumlah Desa
0	Desa Tidak Mengajukan	0.00	0.00	264
1	Desa Mengajukan - Tidak Aktif	0.00	0.00	40
2	Desa Mengajukan - Kurang Aktif	0.07	78.33	56
3	Desa Mengajukan - Aktif	74.67	226.67	21
4	Desa Mengajukan - Sangat Aktif	287.33	435.67	5

Gambar 13. Statistik Hasil Klasterisasi Desa

4.9. Penyebaran

Penelitian ini menggunakan penyebaran aplikasi web sederhana ke server lokal menggunakan kerangka kerja web Python yaitu Flask. Tahap Deployment menggunakan Flask pada penelitian dibuat dengan langkah berikut ini:

4.10. Menyimpan hasil klasterisasi menggunakan algoritma K-Means Clustering

Hasil klasterisasi menggunakan model dengan algoritma k-means clustering yang telah dievaluasi, kemudian dikonversi dan disimpan dalam format csv, sehingga memudahkan untuk dibaca dan dianalisis. CSV adalah format berbasis teks yang dirancang untuk mewakili data dalam bentuk tabel dengan atribut non-spasialnya.

```
1. # simpan hasil klasterisasi dalam format csv
2.
3. df_baru.to_csv('output.csv')
```

Gambar 14. Kode Program Simpan Hasil Klasterisasi Dengan Nama Output.CSV

4.11. Menyimpan hasil visualisasi geospasial berupa peta statis dengan library geopandas

Hasil visualisasi model yang menggunakan library GeoPandas untuk menghasilkan peta statis disimpan dalam format .jpg

```
1. # plot peta statis MMD
2. fig, ax = plt.subplots(figsize = (20,20))
3.
4. keo_boundary_map = gpdkecamatan.boundary.plot(ax=ax,
5. color='Black', linewidth=1.5)
6.
7. for idx, row in gpdkecamatan.iterrows():
8.     plt.annotate(text=idx, xy=(row['geometry'].centroid.x,
9. row['geometry'].centroid.y), horizontalalignment='center',
10. color='black')
11.
12. #plt.savefig('static/images/mmd-kmeans.jpg')
```

Gambar 15. Kode Program Hasil Visualisasi Peta Statis Menggunakan Library Geopandas

4.12. Menyimpan hasil visualisasi geospasial berupa peta dinamis dengan library folium

Hasil visualisasi model yang menggunakan library Folium untuk peta dinamis disimpan dalam format .html. Peta ini memungkinkan interaksi pengguna, seperti zooming dan panning, untuk mengeksplorasi hasil klasterisasi lebih mendalam.

```
4. map = folium.Map(location=[-7.0765, 106.7173],
5. zoom_start=10, tiles='cartodbpositron')
6.
7. for idx, row in df_baru.iterrows():
8.     sim_geo = gpd.GeoSeries(row['geometry'],
9. index=[row['cluster']]).simplify(tolerance=0.001)
10.     sim_geo
11.     geo_j = sim_geo.to_json()
12.     geo_j = folium.GeoJson(
13.         data=geo_j,
14.         style_function = lambda feature: {
15.             'fillColor': get_color(feature),
16.             'fillOpacity': 0.7,
17.             'color': 'black',
18.             'weight': 1
19.         })
20.     folium.Popup(row['DESA_x']).add_to(geo_j)
21.     geo_j.add_to(map)
22.
23. map.save('output.html')
```

Gambar 16. Kode Program Hasil Visualisasi Peta Dinamis Menggunakan Library Folium

4.13. Membuat API (Application Programming Interface) menggunakan Flask

Mengembangkan API menggunakan Flask untuk menghubungkan aplikasi dengan model yang telah disimpan beserta hasil visualisasi (dalam format .jpg dan .html). API ini hanya berfungsi untuk menampilkan hasil klasterisasi dan visualisasi model yang telah disimpan, yang dapat diakses melalui HTTP, tanpa melakukan proses klasterisasi ulang.

```
1.
2. app = Flask(__name__)
3.
4. @app.route('/')
5. def index():
6.     return render_template('index.html')
7.
8. @app.route('/statis')
9. def statis():
```

```

10.     return render_template('static.html')
11.
12. @app.route('/dinamis')
13. def dinamis():
14.     return render_template('dinamis.html')
15.
16. if __name__ == '__main__':
17.     app.run()
18.

```

Gambar 17. Kode Program Untuk Membuat API

4.14. Membuat antarmuka aplikasi menggunakan HTML dan Bootstrap

Mengembangkan antarmuka aplikasi untuk menampilkan hasil klasterisasi dan visualisasi model menggunakan template HTML yang dirender dengan Flask. Desain antarmuka dibuat responsif dengan menggunakan framework CSS Bootstrap untuk memastikan tampilan yang baik di berbagai perangkat. Flask akan merender template HTML dinamis dengan data yang diterima dari API atau model, sehingga pengguna dapat melihat hasil visualisasi secara interaktif

```

1. <!doctype html>
2. <html lang="en">
3. <head>
4.     (< block head >)
5.     <meta charset="utf-8">
6.     <meta name="viewport" content="width=device-width,
7.         initial-scale=1">
8. <title>(< block title >){< endblock >} - MGD Informasi
9.     <link
10.         href="https://cdn.jsdelivr.net/npm/bootstrap@5.3.3/dist/css/
11.             bootstrap.min.css" rel="stylesheet"
12.             integrity="sha384-
13.             GWTM2yjp6P6j1Sv5W4H90F6Rpk6VctnYmDrSp8Iy72BrJk0Jmhy1QW+AL
14.             KxIH"
15.             crossorigin="anonymous">
16.     (< endblock >)
17. </head>
18. <body>
19.
20.
21.
22.
23.
24.
25.
26.
27.
28.
29.
30.
31.
32.
33.
34.
35.
36.
37.
38.
39.
40.
41.
42.
43. <script
44.     src="https://cdn.jsdelivr.net/npm/bootstrap@5.3.3/dist/js/b
45.         ootstrap.bundle.min.js"
46.         integrity="sha384-
47.         YpcerYF0Y31HB6OMNkXc5s5FUVLE2aAA55NDz0xhy90KcidelF1aW7B6j
48.         TeHz"
49.         crossorigin="anonymous"></script>
50. </body>
51. </html>

```

Gambar 18. Kode Program Untuk Membuat Tampilan Antarmuka Aplikasi

4.15. Menjalankan Aplikasi Flask

Jalankan aplikasi Flask di server lokal untuk memastikan API berjalan dengan baik dan bisa diakses oleh pengguna atau sistem lain seperti HTTP dengan cara mengetik perintah python app.py di terminal atau *command prompt* yang sudah diarahkan ke folder tempat file app.py berada.

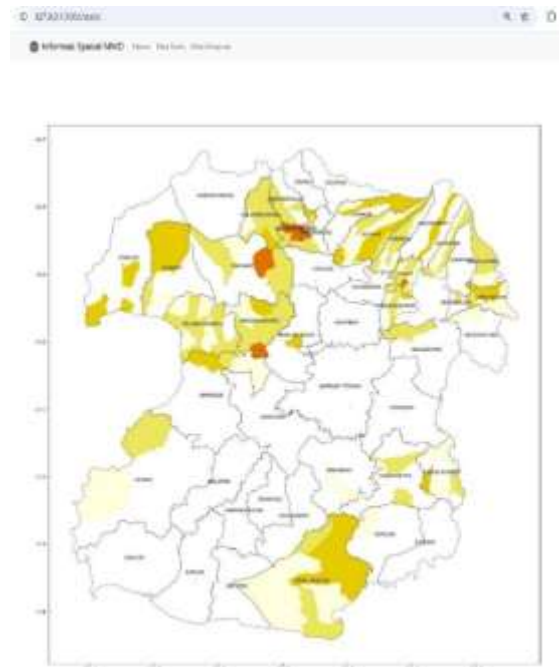
4.16. Akses aplikasi di browser

Setelah server Flask berjalan, aplikasi bisa diakses melalui alamat <http://127.0.0.1:5000/> secara default. Berikut adalah tampilan halaman awal (Beranda) yang menampilkan tampilan berikut



Gambar 19. Halaman Awal Aplikasi MMD

Klik menu statis untuk melihat peta statis yang menampilkan representasi geografis dari data yang telah diproses, dan tampilan akan muncul seperti berikut:



Gambar 20. Halaman Peta Statis

Pada peta yang ditampilkan, warna digunakan untuk membedakan berbagai kelompok atau klaster yang dihasilkan dari pemodelan spasial. Berikut adalah penjelasan mengenai makna dari masing-masing warna pada peta

Tabel 2. Keterangan Warna Peta

No	Warna	Keterangan
1		Klaster desa sangat aktif
2		Klaster desa aktif
3		Klaster desa kurang aktif
4		Klaster desa tidak aktif
5		Klaster desa tidak bekerjasama

Pada halaman peta statis saat *discroll* ke bawah, menampilkan tabel data yang memberikan penjelasan lebih lanjut mengenai hasil klasterisasi.



Cluster	Kode Desa	Nama Desa	Kecamatan	Jumlah Penduduk
A	001001001	001001001	001001001	1000000
B	001001002	001001002	001001001	1000000
C	001001003	001001003	001001001	1000000
D	001001004	001001004	001001001	1000000
E	001001005	001001005	001001001	1000000
F	001001006	001001006	001001001	1000000
G	001001007	001001007	001001001	1000000
H	001001008	001001008	001001001	1000000
I	001001009	001001009	001001001	1000000
J	001001010	001001010	001001001	1000000

Gambar 21. Tampilan Datatable Dataset Hasil Klasterisasi Menggunakan Algoritma K-Means

Menu selanjutnya adalah Peta Dinamis, yang dapat diakses dengan mengklik bagian Navbar Peta Dinamis. Setelah diklik, peta dinamis akan muncul dan menampilkan visualisasi informasi spasial secara interaktif. Dengan mengklik pada area tertentu di peta akan muncul detail data yang terkait dengan wilayah tersebut, seperti kode desa, nama desa, nama kecamatan, statistik pengajuan, dan klasternya



Gambar 22. Tampilan Halaman Peta Dinamis

5. KESIMPULAN DAN SARAN

Berdasarkan pembahasan dalam penelitian ini, dapat disimpulkan bahwa hasil klasterisasi menunjukkan adanya variasi dalam tingkat keaktifan layanan adminduk melalui inovasi MMD, penggunaan Algoritma K-Means Clustering berhasil memberikan gambaran umum mengenai klasterisasi desa, serta penggabungan analisis geospasial dengan hasil klasterisasi yang divisualisasikan dalam peta dapat mengidentifikasi distribusi klaster desa berdasarkan indikator inovasi MMD.

Selain itu juga, terdapat saran untuk mengembangkan penelitian selanjutnya yakni diperlukan program sosialisasi dan pembimbingan teknis untuk mengoptimalkan inovasi MMD,

pemetaan geospasial dapat dikembangkan lebih lanjut karena dapat mengidentifikasi desa yang memerlukan perhatian lebih untuk meningkatkan keaktifan pelayanan, serta perlu adanya peningkatan dalam kualitas dan kuantitas data untuk lebih memaksimalkan penggunaan Algoritma K-Means.

DAFTAR PUSTAKA

- [1] D. Bahtiar *et al.*, “Pemetaan Penduduk Penerima Bantuan Sosial Desa Waru Jaya Menggunakan Algoritma K-Means Clustering,” *Sci. Sacra J. Sains*, vol. 3, no. 2, hal. 29–39, 2023, [Daring]. Tersedia pada: <http://pijarpemikiran.com/index.php/Scientia>
- [2] D. A. Wicaksono dan Y. A. Susetyo, “Clustering Zonasi Daerah Rawan Bencana Alam Di Provinsi Sumatera Barat Menggunakan Algoritma K-Means Dan Library Geopandas,” *J. Indones. Manaj. Inform. dan Komun.*, vol. 4, no. 2, hal. 426–438, 2023, doi: 10.35870/jimik.v4i2.225.
- [3] M. Terra, “Mengetahui Informasi Geospasial Beserta Manfaatnya.” Diakses: 28 November 2024. [Daring]. Tersedia pada: <https://terralogiq.com/informasi-geospasial/>
- [4] M. Saladin, “Mengenal GeoPandas, Library Python Untuk Membantu Menganalisis Data Geospasial.” Diakses: 28 November 2024. [Daring]. Tersedia pada: <https://medium.com/@msfasaladin/mengenal-geopandas-library-python-untuk-membantu-menganalisis-data-geospasial-a8a2fa09438>
- [5] A. Mufid, “Flask: Pengertian, Fungsi, Kelebihan, dan Cara Setup.” Diakses: 7 Desember 2024. [Daring]. Tersedia pada: <https://blog.rumahweb.com/flask-adalah/>
- [6] Anggada Maulana, “Konsep Dasar Data Mining,” *Konsep Data Min.*, vol. 1, hal. 1–16, 2018.
- [7] A. Nur Khomarudin, “Teknik Data Mining : Algoritma K-Means Clustering,” 2003. [Daring]. Tersedia pada: <https://agusnkhom.wordpress.com>
- [8] D. A. I. C. Dewi dan D. A. K. Pramita, “Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali,” *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no. 3, hal. 102–109, 2019, doi: 10.31940/matrix.v9i3.1662.
- [9] M. N. Zhafar *et al.*, “Penerapan Metode Clustering Dengan Algoritma K-Means Untuk Analisa Persebaran Varian Covid-19 (Studi Kasus Kelurahan Antapani Kidul),” *e-Proceeding Eng.*, vol. 2, no. 1, hal. 1042, 2023, [Daring]. Tersedia pada: <http://jurnal.mdp.ac.id>