



Clustering Data Penduduk Desa Menggunakan Algoritma Mean Shift

Tedi Maulani, Elin Haerani*, Fitri Wulandari, Lola Oktavia

Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim, Pekanbaru, Indonesia

Email: ¹1950111744@students.uin-suska.ac.id, ^{2,*}elin.haerani@uin-suska.ac.id, ³fitri_wulandari@yahoo.com, ⁴lola.oktavia@uin-suska.ac.id

Email Penulis Korespondensi: elin.haerani@uin-suska.ac.id

Abstrak—Permasalahan kesejahteraan sosial di Indonesia masih menjadi tantangan serius, termasuk di Provinsi Riau yang memiliki kekayaan sumber daya alam namun belum mampu mewujudkan pemerataan kesejahteraan. Salah satu langkah yang diambil pemerintah untuk mengatasi hal ini adalah dengan memberikan bantuan sosial kepada masyarakat, namun penentuan penerima bantuan masih menemui kendala. Penelitian ini bertujuan untuk mengelompokkan masyarakat Desa Bina Baru menggunakan algoritma Mean Shift agar penyaluran bantuan sosial lebih tepat sasaran. Evaluasi dilakukan menggunakan Silhouette Score untuk mengukur kualitas pengelompokan. Hasil penelitian menunjukkan bahwa pengelompokan optimal terjadi pada quantile 0.9 dengan Silhouette Score tertinggi sebesar 0.5747, menghasilkan sembilan cluster dengan karakteristik sosial ekonomi berbeda. Berdasarkan hasil analisis, dapat disimpulkan bahwa cluster 2, cluster 1, cluster 5, dan cluster 6/7 merupakan kelompok yang paling layak diprioritaskan untuk menerima bantuan pemerintah. Penentuan prioritas ini sangat berguna dalam penyaluran bantuan yang lebih tepat sasaran dan berbasis data. Pendekatan ini memberikan rekomendasi yang lebih akurat bagi pemerintah dalam menetapkan kebijakan distribusi bantuan serta mendukung perencanaan pengentasan kemiskinan yang berkelanjutan di wilayah Desa Bina Baru.

Kata Kunci: Clustering; Data Mining; Mean Shift; Penduduk; Silhouette Score

Abstract—Social welfare remains a serious challenge in Indonesia, including in Riau Province, which, despite its abundant natural resources, still struggles with unequal distribution of welfare. One of the government's efforts to address this issue is through social assistance programs. However, identifying the right beneficiaries remains problematic. This study aims to cluster residents of Desa Bina Baru using the Mean Shift algorithm to support more targeted social aid distribution. The clustering results were evaluated using the Silhouette Score to measure their quality. The optimal clustering was achieved at a quantile of 0.9, with the highest Silhouette Score of 0.5747, producing nine clusters with varying socioeconomic characteristics. Based on the analysis, clusters 2, 1, 5, and 6/7 were identified as the most eligible groups to receive government aid due to economic pressure, high number of dependents, and inadequate housing conditions. This prioritization is crucial for more accurate, data-driven distribution of aid and provides valuable insights to support sustainable poverty alleviation strategies in Desa Bina Baru.

Keywords: Clustering; Data Mining; Mean Shift; Population; Silhouette Score

1. PENDAHULUAN

Sejak masa awal kemerdekaan Indonesia, pemerintah telah memberikan perhatian khusus dan mendalam untuk menciptakan kesejahteraan di masyarakat. Namun hingga saat ini, masyarakat Indonesia belum merasakan kesejahteraan secara keseluruhan, sehingga hal tersebut menjadi pekerjaan rumah pemerintah yang sulit dicapai. Salah satu penyebab sulitnya mencapai kesejahteraan bagi masyarakat ialah masalah yang sulit untuk diberantas, seperti rendahnya tingkat akses air bersih, permukiman warga yang padat dan kumuh, rendahnya kualitas pendidikan, dan tingkat pendapatan masyarakat yang masih jauh dari layak (Lisnawati Sopiah & Ria Haryatiningsih, 2023).

Isu kesejahteraan sosial selalu menjadi pembahasan utama di setiap negara di dunia. Penanggulangan permasalahan ini yang menjadi persoalan pemerintah selanjutnya akan menciptakan kesejahteraan sosial yang merupakan tujuan akhir dari suatu negara (Sinurat, 2023). Permasalahan ini tidak tersebar secara merata di Indonesia. Terdapat beberapa provinsi dengan angka yang rendah dan beberapa provinsi lainnya yang memiliki tingkat lebih tinggi. Wilayah Kawasan Barat Indonesia (KBI) mendominasi tingkat penduduk miskin sebesar 74,45%, terutama di wilayah Sumatera dan Jawa. Hal ini terbilang wajar karena sebagian besar penduduk Indonesia berada di kedua pulau tersebut. Salah satu provinsi yang termasuk di dalamnya adalah Provinsi Riau (Sangadah et al., 2020). Provinsi Riau merupakan salah satu provinsi yang memiliki kekayaan khusus, yang terbilang unggul dibandingkan provinsi lainnya di Pulau Sumatera, dengan jumlah penduduk yang relatif tinggi dan kekayaan sumber daya alam yang berlimpah. Kekayaan sumber daya alam yang ada seharusnya dapat mendorong perekonomian masyarakat Riau untuk menjadi lebih baik, sehingga dapat mengurangi dampak kesenjangan sosial (Sari et al., 2022). Namun, walaupun dengan sumber daya alam yang melimpah, Provinsi Riau masih menghadapi tantangan dalam mengatasi ketidakmerataan kesejahteraan di masyarakat (Aldisun et al., 2022).

Dalam menangani tingkat kesejahteraan di Provinsi Riau, pemerintah memberikan bantuan berupa tunai maupun non-tunai kepada setiap keluarga, agar mendorong minat belanja dan meningkatkan perekonomian di Provinsi Riau. Namun, masih ada kendala dalam penentuan siapa saja pihak yang tepat dan berhak untuk menerima bantuan langsung dari pemerintah. Untuk itu, diperlukan pengelompokan berdasarkan kriteria tertentu untuk memudahkan pengelompokan keluarga tersebut ke dalam beberapa kategori. Salah satu cara yang digunakan adalah pengelompokan data menggunakan algoritma Mean Shift (Rianti et al., 2024). Mean Shift adalah sebuah metode algoritma non-parametrik. Algoritma ini memiliki kemampuan untuk menentukan jumlah cluster secara adaptif, berbeda dengan pendekatan clustering lainnya yang biasanya memerlukan jumlah cluster ditentukan terlebih dahulu (Zhao et al., 2021).

Cara kerjanya adalah dengan menggunakan jendela geser berbentuk lingkaran untuk melakukan pengelompokan data (Rizuan et al., 2023). Algoritma Mean Shift digunakan untuk mengenakan suatu kumpulan warna pada area tertentu yang memiliki kerapatan data tinggi (Purba et al., 2019).

Dalam beberapa penelitian sebelumnya menggunakan algoritma Mean Shift Clustering digunakan untuk mengelompokkan wilayah berdasarkan data pengelolaan sampah di kabupaten/kota. Dengan bandwidth 1.5, terbentuk dua kluster dengan Silhouette Score sebesar 0.649 yang menunjukkan kualitas pengelompokan yang cukup baik, yaitu kluster dengan sampah terkelola rendah dan kluster dengan pengelolaan sampah yang tinggi (Lidya Musaffak et al., 2025). Kemudian pada penelitian yang menerapkan algoritma Mean Shift untuk mengkluster distribusi listrik di Indonesia guna meningkatkan efisiensi energi, khususnya di wilayah terpencil dengan bandwidth optimal 0.5, terbentuk tiga kluster di setiap wilayah utama, mencerminkan kesamaan infrastruktur, kebutuhan energi, dan sektor ekonomi dengan hasil ini menunjukkan fleksibilitas Mean Shift dalam menangani data kompleks tanpa menentukan jumlah kluster sebelumnya, serta mendukung perencanaan distribusi listrik yang lebih efisien dan berkelanjutan (Utari et al., 2024). Penelitian selanjutnya menggunakan algoritma Mean Shift dalam segmentasi hutan berdasarkan citra menunjukkan bahwa algoritma Mean Shift dengan pendekatan baru untuk estimasi bandwidth dan modul ekstraksi warna dengan bandwidth optimal 25.38 dan 30 iterasi, skenario terbaik mencapai akurasi 90.75%. Algoritma ini efisien secara waktu meskipun dijalankan pada perangkat non-high-end (Andrei & Grigore, 2023). Pada penelitian yang membandingkan K-Means, Mean Shift, dan DBSCAN untuk mengelompokkan ancaman jaringan pada instansi kesehatan. Hasil menunjukkan Mean Shift sebagai algoritma terbaik dengan nilai Silhouette Index tertinggi sebesar 0,5305, sehingga paling efektif dalam mengidentifikasi pola ancaman secara akurat (Cinderatama et al., 2022).

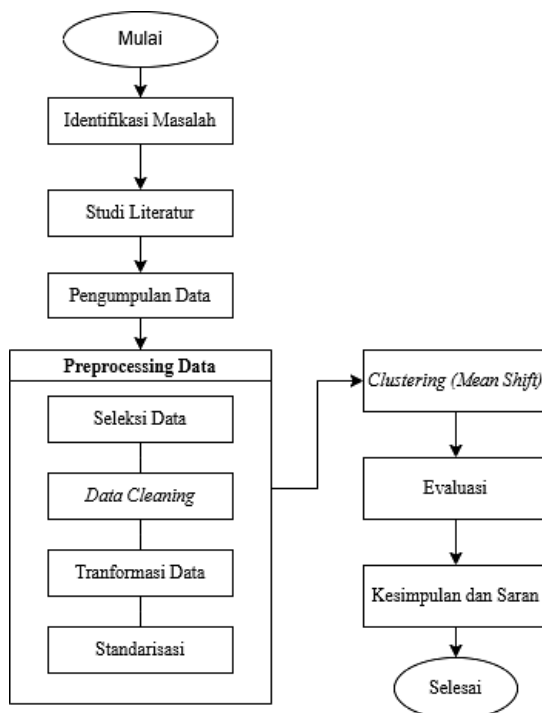
Untuk menganalisis hasil kinerja dari algoritma Mean Shift, maka akan dilakukan evaluasi dengan Silhouette Score (Punhani et al., 2022). Salah satu pengukuran untuk mengevaluasi kualitas teknik clustering adalah Silhouette Score (Ling & Weiling, 2025). Silhouette Score adalah proses evaluasi yang digunakan untuk mengukur seberapa baik suatu algoritma clustering, yang dalam penelitian ini adalah algoritma Mean Shift, di mana Silhouette Score mampu memisahkan data menjadi kelompok-kelompok yang berbeda (Tipwong et al., 2025).

Dari pemaparan di atas, penulis bertujuan untuk mengelompokkan data seluruh masyarakat di wilayah Desa Bina Baru untuk mendapatkan data keluarga yang layak mendapatkan bantuan dari pemerintah menggunakan beberapa atribut, mulai dari status kepemilikan rumah, jenis lantai, jenis pendapatan, sumber air, dan beberapa kriteria lainnya.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan penelitian adalah rencana atau kerangka kerja yang menjelaskan bagaimana suatu penelitian akan dilakukan. Desain penelitian menampilkan tahapan dari penelitian yang akan dikerjakan secara sistematis dan logis, untuk mencapai tujuan penelitian. Pada Gambar 1 dibawah ini, akan tampilan tahapan yang akan dilakukan pada penelitian ini



Gambar 1. Tahapan Penelitian



2.2 Identifikasi Masalah

Tahap identifikasi masalah adalah langkah awal yang penting dilakukan untuk memahami dan mencari solusi terhadap permasalahan. Fokus masalah dalam penelitian ini adalah menentukan cluster untuk mengelompokkan data seluruh masyarakat di wilayah Desa Bina baru untuk mendapatkan data keluarga yang layak mendapatkan bantuan dari pemerintah menggunakan algoritma Mean Shift.

2.3 Studi Literatur

Studi literatur merupakan proses pencarian referensi teori yang berhubungan dengan penelitian yang dilakukan. Studi literatur ini dilakukan untuk mengetahui tentang Data Mining, Clustering, Mean Shift, dan Silhouette Score, yang dapat bersumber dari jurnal, artikel, prosiding, dan e-book dengan kata kunci 'Data Mining', 'Clustering', 'Mean Shift', dan 'Silhouette Score'.

2.4 Pengumpulan Data

Pengumpulan data adalah proses sistematis untuk mengumpulkan, mengorganisir, dan menyimpan informasi yang diperlukan untuk penelitian, analisis, atau pengambilan keputusan. Proses ini sangat penting dalam melakukan penelitian karena data yang akurat dan relevan menjadi dasar untuk mencapai kesimpulan yang valid. Pengumpulan data bertujuan untuk merancang perencanaan dan strategi penelitian, melakukan analisis dan evaluasi terhadap hasil penelitian, dan pengambilan keputusan terhadap hasil penelitian. Dataset yang digunakan dalam penelitian ini berasal dari data SDGs (Sustainable Development Goals) Desa Bina Baru tahun 2022, khususnya pada bagian data keluarga. Data ini dikumpulkan oleh pihak desa sebagai bentuk pemetaan sosial dan ekonomi masyarakat. Data tersebut memuat berbagai informasi penting yang mencerminkan kondisi sosial ekonomi setiap keluarga. Tabel 1 berikut ini adalah contoh bentuk data yang digunakan dalam penelitian:

Tabel 1. Contoh Bentuk Data

No	Penghasilan Anggota Keluarga	Sumber Penghasilan	...	Kondisi Rumah
1	2500000	2	Buruh Tani	... Tidak Kumuh
2	8000000	3	Pemilik Kebun Sawit	... Tidak Kumuh
3	1500000	2	Buruh Tani	... Tidak Kumuh
4	3500000	6	Buruh Tani	... Tidak Kumuh
5	3000000	3	Buruh Tani	... Tidak Kumuh
...
1005	3000000	3	Buruh Tani	... Tidak Kumuh

2.5 Data Preprocessing

Data pre-processing merupakan tahapan yang terjadi setelah melakukan pengumpulan data yang akan digunakan dalam penelitian. Pada tahap ini, data diselidiki dan dianalisis secara menyeluruh untuk menentukan masalah yang terjadi, sebelum mengambil tindakan atau mengambil keputusan (Maharana et al., 2022). Adapun proses yang terdapat pada tahap preprocessing adalah sebagai berikut:

2.5.1 Data Selection

Data Selection atau seleksi data merupakan tahapan yang diperlukan untuk menyeleksi atribut yang akan digunakan dalam penelitian (Setiawan, 2023). Dalam penelitian ini, atribut yang akan digunakan adalah Penghasilan, Pengeluaran, Anggota Keluarga, Sumber Penghasilan, Dinding Rumah, Lantai Rumah, Status Rumah, Atap Rumah, Tempat Buang Air Besar, MCK, Sumber Air Minum, Penerangan Utama, Bahan Bakar Memasak, Pembuangan Sampah, Pembuangan Limbah Cair, dan Kondisi Rumah. Proses seleksi ini juga menghapus kolom-kolom yang tidak memberikan kontribusi informatif terhadap tujuan analisis, seperti kolom *nama*, yang bersifat identitas dan tidak relevan untuk proses prediksi.

2.5.2 Data Cleaning

Tahapan ini merupakan tahapan pembersihan data dari data yang missing value. Tahapan ini dilakukan untuk mendapatkan hasil yang valid dari hasil clustering penelitian (Aldoseri et al., 2023). Salah satu permasalahan yang ditemukan dalam data adalah adanya missing value atau nilai kosong, terutama pada kolom "*Pembuangan Sampah*" dan "*Pembuangan Limbah Cair*". Untuk menangani hal ini, dilakukan imputasi menggunakan metode modus, yaitu menggantikan nilai kosong dengan nilai yang paling sering muncul pada masing-masing kolom tersebut. Setelah proses *cleaning* selesai, jumlah fitur dalam dataset menjadi 16 kolom, dan jumlah data yang tersedia menyusut menjadi 1.000 baris yang dinilai layak untuk dianalisis lebih lanjut.

2.5.3 Data Transformation

Tahapan ini merupakan merubah data kebentuk yang diperlukan sesuai dengan pembahasan penelitian. Data yang telah di dibersihkan dari data-data yang tidak diperlukan kemudian akan berlanjut ke proses pengubahan. Proses perubahan data yaitu mengubah data ke dalam bentuk angka (Fan et al., 2021). Transformasi ini merupakan tahap penting agar model dapat memahami dan memproses data dengan benar, sehingga hasil prediksi atau klasifikasi menjadi lebih akurat

dan bermakna. Rincian kode transformasi yang digunakan dapat dilihat pada Tabel 2, sedangkan hasil transformasi terhadap data disajikan pada Tabel 3.

Tabel 2. Kode Transformasi Data

No	Nama Variabel/Atribut	Keterangan
1	Penghasilan keluarga perbulan	Gaji perbulan
2	Pengeluaran keluarga perbulan	Berapa banyak pengeluaran perbulan
3	Anggota keluarga	Berapa orang menjadi tanggungan oleh Kepala keluarga 1. Buruh tani. 2. Pemilik Kebun Sawit. 3. Pedagang.
4	Sumber penghasilan	4. Karyawan. 5. PNS. 6. Guru Bantu. 7. Peternakan. 8. Bengkel. 9. Walet. 10. Satpam.
5	Jenis dinding rumah	1. Semen. 2. Kayu.
6	Jenis Lantai Rumah	1. Semen. 2. Keramik. 3. Kayu. 4. Ubin. 5. Marmer.
7	Status Penguasaan Rumah	1. Sendiri. 2. Bebas Sewa. 3. Kontrak. 4. Dinas.
8	Jenis Atap Rumah	1. Genteng. 2. Seng. 3. Ijuk.
9	Fasilitas Tempat Buang Air Besar	1. Jamban Sendiri. 2. Jamban umum.
10	Fasilitas MCK	1. Sendiri. 2. MCK umum.
11	Sumber Air Minum	1. Sumber air. 2. Mata air/sumur.
12	Sumber Penerangan Utama	1. Listrik PLN. 2. Lilin.
13	Jenis Bahan Bakar untuk Memasak	1. Gas kota/LPG. 2. Minyak Tanah.
14	Tempat Pembuangan Sampah	1. Dibuang. 2. Tempat Sampah.
15	Tempat Pembuangan Limbah Cair	1. Lubang Ditanah. 2. Tangki.
16	Secara Keseluruhan Kondisi Rumah	1. Kumuh. 2. Tidak Kumuh.

Tabel 3. Transformasi Data

No	1	2	3	4	...	16
1	2500000	2000000	2	1	...	2
2	8000000	4000000	3	2	...	2
3	1500000	1000000	2	1	...	2
4	3500000	3000000	6	1	...	2
5	3000000	2500000	3	1	...	2

2.5.4 Standarisasi

Standarisasi adalah salah satu teknik preprocessing data yang digunakan untuk menskalakan nilai fitur agar memiliki distribusi dengan rata-rata (mean) sebesar 0 dan standar deviasi sebesar 1 (Zamri et al., 2022). Teknik ini penting terutama ketika fitur memiliki skala yang berbeda, karena dapat mempengaruhi performa algoritma.

2.6 Data Mining

Data mining merupakan proses mengubah kumpulan data dalam jumlah besar menjadi pengetahuan yang bermakna dengan memanfaatkan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin (machine learning)(Surbakti, 2021; Wang et al., 2023). Salah satu metode yang digunakan dalam proses ini adalah *Mean Shift Clustering*, yaitu algoritma non-parametrik yang mampu mengelompokkan data tanpa perlu menentukan jumlah cluster di awal. Dengan pendekatan ini, data dapat dikelompokkan secara adaptif berdasarkan distribusi dan kepadatan data, sehingga sangat efektif dalam menemukan struktur alami dalam dataset(Farhan Nugraha et al., 2024).

2.7 Clustering Mean Shift

Algoritma Mean Shift adalah sebuah metode non-parametrik(Ghassabeh & Rudzicz, 2021). Algoritma ini dapat digunakan salah satunya pada pengelompokan data(Ren et al., 2025). Algoritma Mean Shift merupakan suatu teknik pengelompokan data yang memfokuskan pada penemuan titik pusat atau "nodes" melalui sebaran data yang menggunakan metode pergeseran jendela untuk mengidentifikasi wilayah-wilayah dengan tingkat kepadatan data yang tinggi(Aranguren et al., 2025). Mean Shift merupakan metode yang dapat digunakan untuk membagi dataset menjadi beberapa kelompok (cluster) dengan cara memberikan label pada setiap titik data. Label tersebut ditentukan berdasarkan titik konvergen, yaitu titik akhir yang dicapai setelah beberapa iterasi menggunakan persamaan pembaruan dan menggeser titik data ke arah rata-rata tetangganya dalam radius tertentu yang disebut bandwidth. Bandwidth menentukan seberapa luas area pencarian titik-titik terdekat. Untuk menentukan bandwidth secara otomatis, digunakan parameter quantile menentukan persentase jarak antar titik yang dipakai dalam estimasi. Quantile adalah nilai persentase (antara 0 dan 1) yang digunakan untuk menghitung bandwidth berdasarkan distribusi jarak antar titik dalam data. Titik-titik yang konvergen ke lokasi yang sama akan dikategorikan dalam satu cluster(Cariou et al., 2022; Chakraborty et al., 2021).yang di rumuskan pada persamaan berikut(Cariou et al., 2022):

$$m(x) = \frac{\sum_{i=1}^n K(x_i - x) \cdot x_i}{\sum_{i=1}^n K(x_i - x)} \quad (1)$$

Mean Shift adalah metode clustering yang bekerja dengan cara memindahkan titik data menuju daerah dengan densitas data yang lebih tinggi. Dalam prosesnya, posisi saat ini ditunjukkan oleh x , sedangkan x_i merepresentasikan data point ke- i dalam dataset. Mean Shift menggunakan fungsi kernel K , seperti Gaussian kernel, untuk menghitung kontribusi tiap data point terhadap pusat baru. Hasil dari perhitungan ini adalah vektor Mean Shift $m(x)$, yang menunjukkan arah dan besarnya pergeseran posisi x menuju pusat densitas maksimum di sekitarnya.

2.7 Evaluasi

Pada penelitian ini, pengujian hasil dilakukan dengan menggunakan Silhouette Score yang merupakan salah satu metrik validasi internal yang digunakan untuk mengevaluasi kualitas hasil clustering. Metrik ini mengukur sejauh mana objek-objek dalam satu cluster memiliki kemiripan satu sama lain (*compactness*), sekaligus memastikan bahwa objek tersebut cukup berbeda atau terpisah dari cluster lainnya (*separation*). Nilai silhouette berkisar antara -1 hingga 1, di mana nilai mendekati 1 menunjukkan bahwa objek berada dalam cluster yang tepat (Althobaiti et al., 2023). Pengujian Silhouette Score juga dapat membantu dalam mengevaluasi performa algoritma clustering yang digunakan, memungkinkan penyesuaian parameter atau pemilihan metode clustering yang lebih baik jika hasilnya tidak memuaskan (Hendrastuty, 2024). Interpretasi dari nilai Silhouette Coefficient secara umum dapat dilihat pada Tabel 4.

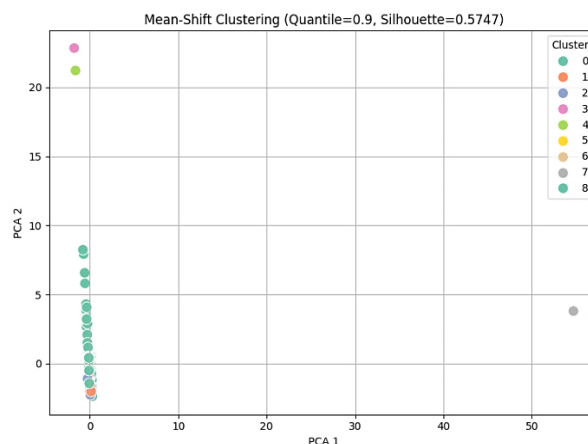
Tabel 4. Interpretasi Silhouette Coefficient

<i>Silhouette Coefficient</i>	Interpretasi
0.71 – 1.00	Struktur yang dihasilkan kuat
0.51 – 0.70	Struktur yang dihasilkan baik
0.26 – 0.50	Struktur yang dihasilkan lemah
≤ 0.25	Tidak terstruktur

3. HASIL DAN PEMBAHASAN

3.1 Penerapan Algoritma Clustering Meant Shift

Dalam proses clustering data ekonomi keluarga penduduk, digunakan algoritma Mean Shift yang tidak memerlukan penentuan jumlah cluster di awal. Data terlebih dahulu dinormalisasi menggunakan StandardScaler untuk menghilangkan bias akibat perbedaan skala antar fitur. Kemudian, dilakukan eksplorasi nilai parameter quantile dari 0.1 hingga 0.9 untuk mencari bandwidth terbaik yang memberikan performa clustering optimal, Gambar 2 menunjukkan visualisasi dari hasil clustering yang telah dilakukan.



Gambar 2. Visualisasi Hasil Clustering

Untuk memudahkan interpretasi, data direduksi dimensinya menggunakan PCA (Principal Component Analysis) menjadi dua komponen utama (PCA satu dan PCA dua), kemudian divisualisasikan dalam scatter plot. Dari hasil visualisasi, terlihat bahwa sebagian besar data terkonsentrasi di sekitar nilai PCA satu = 0

Hal ini menunjukkan adanya kelompok-kelompok penduduk dengan kondisi ekonomi yang sangat berbeda dari mayoritas. Proses clustering ini sangat bermanfaat dalam mengidentifikasi segmentasi ekonomi keluarga, sehingga dapat digunakan untuk merancang intervensi kebijakan yang lebih tepat sasaran.

3.2 Evaluasi Mean Shift Clustering dengan Silhouette Score

Pengujian performa algoritma Mean Shift clustering dilakukan dengan menggunakan metrik Silhouette Score untuk mengevaluasi kualitas pembentukan cluster. Skor ini menunjukkan seberapa baik suatu data berada dalam clusternya



sendiri dibandingkan dengan cluster lainnya. Semakin tinggi nilai Silhouette Score, semakin baik pemisahan antar cluster dan semakin baik kualitas clustering-nya. Tabel 5 menyajikan perbandingan hasil clustering dan nilai evaluasinya :

Tabel 5. Perbandingan Hasil Evaluasi

Jumlah Cluster	Silhouette Score
51	0.2989
33	0.3148
29	0.3461
25	0.3699
21	0.3734
17	0.4773
15	0.4829
13	0.4861
9	0.5747

Berdasarkan Tabel 4, terlihat bahwa semakin sedikit jumlah cluster, nilai Silhouette Score cenderung meningkat. Saat jumlah cluster masih banyak, seperti 51 atau 33, nilai skor masih rendah (0.2989 dan 0.3148), yang menandakan bahwa pembagian cluster masih kurang optimal atau banyak cluster yang tumpang tindih. Namun, saat jumlah cluster semakin sedikit, nilai Silhouette Score mengalami peningkatan yang cukup signifikan. Nilai tertinggi tercapai saat jumlah cluster adalah 9, yaitu 0.5747, yang menunjukkan bahwa pembagian cluster pada titik ini paling optimal dibandingkan jumlah lainnya. Dengan demikian, jumlah cluster 9 dapat dipertimbangkan sebagai jumlah yang paling ideal untuk data ini, karena memberikan pemisahan antar cluster yang paling baik menurut metrik Silhouette Score.

3.3 Hasil Analisis Centroid menggunakan Mean Shift

Hasil clustering menggunakan algoritma Mean Shift menghasilkan sembilan kelompok (cluster) dengan karakteristik sosial ekonomi dan kondisi rumah tangga yang berbeda seperti yang terlihat pada tabel 6.

Tabel 6. Hasil Clustering

Cluster	0	1	2	3	4	5	6	7	8
penghasilan	3436164, 1	2083333, 33	200000 0	800000 00	800000 00	350000 0	300000 0	300000 0	100000 0
pengeluaran	2483487, 18	2011111, 11	300000 0	600000 00	600000 00	300000 0	250000 0	250000 0	500000
anggota keluarga	3,130256 41	4,5	6	4	4	5	6	6	1
sumber penghasilan	1,566153 85	1,388888 89	1	2	2	1	8	8	1
dinding rumah	1,132307 69	1,277777 78	1,5	1	1	1	1	1	1
lantai rumag	1,514871 79	1,333333 33	2	1	1	1	1	1	1
satatus rumah	1,165128 21	1,222222 22	1	1	1	1	1	1	1
atap rumah	1,593846 15	1,5	1	2	2	1	2	2	1
tempat buan air besar	1	1	1	1	1	1	1	1	1
mck	1	1	1	1	1	1	1	1	1
sumber air minum	1,018461 54	1,833333 33	2	1	1	1	2	2	1
penerangan utama	1	1	1	1	1	1	1	1	1
bahan bakar memasak	1	2	2	1	1	2	2	2	1
pembuangan sampah	1,072820 51	1	1	2	1	2	1	1	1
pembuangan limbah cair	1,033846 15	1	1	2	1	2	1	1	1
kondisi rumah	2	2	1	2	2	2	1	1	1

Hasil clustering menggunakan algoritma Mean Shift menghasilkan sembilan kelompok (cluster) dengan karakteristik sosial ekonomi dan kondisi rumah tangga yang berbeda. Analisis ini bertujuan untuk mengidentifikasi



keluarga-keluarga yang layak menerima bantuan pemerintah berdasarkan sejumlah atribut seperti penghasilan, pengeluaran, jumlah anggota keluarga, sumber penghasilan, kondisi rumah, dan fasilitas dasar lainnya.

Cluster 0 merupakan kelompok dengan rata-rata penghasilan sekitar Rp 3,4 juta dan pengeluaran Rp 2,4 juta. Rata-rata anggota keluarga sebanyak 3 orang, dengan sumber penghasilan didominasi oleh buruh tani atau pemilik kebun sawit. Kondisi rumah dalam kelompok ini cenderung layak, terlihat dari penggunaan dinding semen, atap genteng, serta fasilitas MCK dan sumber air yang memadai. Selain itu, rumah-rumah dalam kelompok ini juga tidak tergolong kumuh. Oleh karena itu, keluarga dalam cluster ini dapat dikategorikan sebagai kelompok masyarakat cukup mampu dan bukan prioritas utama penerima bantuan.

Cluster 1 memiliki penghasilan lebih rendah, yaitu sekitar Rp 2 juta, namun dengan jumlah tanggungan lebih besar, yakni sekitar 4–5 orang per keluarga. Sumber penghasilan masih dari sektor informal, dan terdapat indikasi pemanfaatan sumber air dan bahan bakar memasak yang kurang layak. Meski kondisi rumah cukup baik dan tidak kumuh, beban tanggungan yang besar serta keterbatasan dalam fasilitas dasar menjadikan kelompok ini sebagai kandidat yang layak untuk diprioritaskan menerima bantuan sosial.

Berbeda dengan dua kelompok sebelumnya, cluster 2 menunjukkan kondisi yang lebih memprihatinkan. Dengan penghasilan sebesar Rp 2 juta dan pengeluaran mencapai Rp 3 juta, kelompok ini mengalami defisit ekonomi. Ditambah lagi, jumlah anggota keluarga yang besar (6 orang) serta kondisi rumah yang tergolong kumuh menjadikan kelompok ini sebagai prioritas tinggi untuk menerima bantuan. Kondisi sanitasi dan sumber daya energi juga menunjukkan keterbatasan.

Cluster 3 dan 4 memiliki karakteristik yang hampir serupa, yaitu dengan penghasilan dan pengeluaran yang sangat tinggi (masing-masing Rp 8 juta dan Rp 6 juta). Anggota keluarga berkisar 4 orang, dan semua indikator fasilitas dasar, kondisi rumah, serta sanitasi berada pada tingkat yang sangat baik. Rumah-rumah dalam kelompok ini jelas tidak kumuh. Oleh karena itu, keluarga dalam cluster ini tergolong sebagai kelompok masyarakat sangat mampu dan tidak layak untuk menerima bantuan pemerintah.

Cluster 5 memiliki penghasilan dan pengeluaran yang cukup seimbang di kisaran Rp 3,5 juta dan Rp 3 juta. Meski demikian, kondisi sanitasi dan pengelolaan limbah rumah tangga masih tergolong rendah, dan rumah juga termasuk kumuh. Maka, kelompok ini tetap perlu dipertimbangkan dalam skema bantuan, terutama untuk peningkatan kualitas lingkungan dan sanitasi rumah tangga.

Sementara itu, cluster 6 dan 7 menunjukkan profil serupa: penghasilan dan pengeluaran yang relatif menengah, anggota keluarga besar (6 orang), serta sumber penghasilan dari sektor informal seperti bengkel dan peternakan. Meski beberapa indikator fasilitas cukup memadai, masih ditemukan penggunaan minyak tanah dan air dari sumber yang kurang layak. Keduanya menunjukkan potensi sebagai kelompok yang memerlukan perhatian dalam bentuk bantuan peningkatan fasilitas dasar.

Cluster 8 merupakan kelompok dengan penghasilan dan pengeluaran paling rendah (masing-masing Rp 1 juta dan Rp 500 ribu), namun hanya memiliki satu anggota keluarga dan seluruh fasilitas rumah tergolong layak serta rumah tidak kumuh. Oleh karena itu, meskipun secara ekonomi berada di bawah, rendahnya beban tanggungan dan layaknya fasilitas menjadikan cluster ini bukan prioritas utama dalam penerimaan bantuan.

Berdasarkan hasil analisis tersebut, dapat disimpulkan bahwa cluster 2, cluster 1, cluster 5, dan cluster 6/7 merupakan kelompok yang paling layak diprioritaskan untuk menerima bantuan pemerintah karena mengindikasikan adanya tekanan ekonomi, jumlah tanggungan besar, serta kondisi rumah yang tidak memadai. Penentuan prioritas ini sangat berguna dalam penyaluran bantuan yang lebih tepat sasaran, dan berbasis data.

4. KESIMPULAN

Berdasarkan proses analisis yang telah dilakukan, dapat disimpulkan bahwa algoritma Mean-Shift berhasil mengelompokkan data penduduk Desa Bina Baru ke dalam sembilan klaster dengan karakteristik sosial ekonomi dan kondisi rumah tangga yang berbeda-beda. Pengelompokan optimal diperoleh pada nilai quantile sebesar 0.9 dengan Silhouette Score tertinggi sebesar 0.5747, yang menunjukkan kualitas pengelompokan cukup baik dalam membedakan karakteristik antar individu. Analisis centroid masing-masing klaster mengungkapkan adanya kelompok masyarakat yang sangat rentan secara ekonomi dan lingkungan, seperti pada cluster 2 yang memiliki penghasilan rendah, pengeluaran tinggi, jumlah tanggungan besar, serta kondisi rumah kumuh dan fasilitas terbatas. Cluster 1, 5, 6, dan 7 juga mencerminkan keluarga dengan beban ekonomi menengah ke bawah, fasilitas sanitasi kurang memadai, dan kondisi sosial yang membutuhkan perhatian, sehingga dapat menjadi prioritas utama dalam penyaluran bantuan sosial. Sebaliknya, cluster 3 dan 4 menunjukkan karakteristik keluarga dengan pendapatan tinggi dan fasilitas tempat tinggal yang baik, sementara cluster 0 dan 8 menggambarkan kondisi ekonomi relatif stabil meskipun berada pada tingkat penghasilan menengah atau rendah. Kesimpulan utama dari penelitian ini adalah bahwa algoritma Mean Shift efektif dalam mengelompokkan data sosial ekonomi masyarakat berdasarkan atribut penting, serta membuka peluang untuk pelaksanaan program sosial yang lebih terukur, tepat sasaran, dan berbasis data.



REFERENCES

- Aldisun, F., Ukas, U., & Zuhdi Arman, Z. A. (2022). Analisis Yuridis Pengelolaan Sumber Daya Alam Kelautan-Perikanan Berdasarkan Undang-Undang Nomor 32 Tahun 2014 Di Kepulauan Riau. *Jurnal Cahaya Keadilan*, 10(1), 1–12. <https://doi.org/10.33884/jck.v10i1.5511>
- Aldoseri, A., Al-Khalifa, K. N., & Hamouda, A. M. (2023). Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges. *Applied Sciences*, 13(12), 7082. <https://doi.org/10.3390/app13127082>
- Althobaiti, K., Wolters, M. K., Alsufyani, N., & Vaniea, K. (2023). Using Clustering Algorithms to Automatically Identify Phishing Campaigns. *IEEE Access*, 11, 96502–96513. <https://doi.org/10.1109/ACCESS.2023.3310810>
- Andrei, A.-T., & Grigore, O. (2023). Mean Shift Clustering with Bandwidth Estimation and Color Extraction Module Used in Forest Segmentation. 2023 13th International Symposium on Advanced Topics in Electrical Engineering (ATEE), 1–6. <https://doi.org/10.1109/ATEE58038.2023.10108106>
- Aranguren, I., Valdivia, A., & Navarro, J. (2025). Mean-Shift Clustering for Failure Detection in Quadcopter Unmanned Aerial Vehicles. In D. Oliva, A. Valdivia, S. J. Mousavirad, & K. Kalita (Eds.), *Advances in Optimization Algorithms for Multidisciplinary Engineering Applications: From Classical Methods to AI-Enhanced Solutions* (pp. 785–806). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-78440-8_35
- Cariou, C., Le Moan, S., & Chehdi, K. (2022). A Novel Mean-Shift Algorithm for Data Clustering. *IEEE Access*, 10, 14575–14585. <https://doi.org/10.1109/ACCESS.2022.3147951>
- Chakraborty, S., Paul, D., & Das, S. (2021). Automated Clustering of High-dimensional Data with a Feature Weighted Mean Shift Algorithm. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(8), 6930–6938. <https://doi.org/10.1609/aaai.v35i8.16854>
- Cinderatama, T. A., Alhamri, R. Z., & Yunhasnawa, Y. (2022). Implementasi Metode K-Means, DbSCAN, Dan Meanshift Untuk Analisis Jenis Ancaman Jaringan Pada Intrusion Detection System. *INOVTEK Polbeng - Seri Informatika*, 7(1), 169. <https://doi.org/10.35314/isi.v7i1.2336>
- Fan, C., Chen, M., Wang, X., Wang, J., & Huang, B. (2021). A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data. *Frontiers in Energy Research*, 9. <https://doi.org/10.3389/fenrg.2021.652801>
- Farhan Nugraha, M., Martano, M., & Hayati, U. (2024). Clustering Data Indonesian Food Delivery Menggunakan Metode K-Means Pada Gofood Product List. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 3484–3492. <https://doi.org/10.36040/jati.v8i3.9727>
- Ghassabeh, Y. A., & Rudzicz, F. (2021). Modified Subspace Constrained Mean Shift Algorithm. *Journal of Classification*, 38(1), 27–43. <https://doi.org/10.1007/s00357-019-09353-1>
- Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika Dan Ilmu Komputer (JIMA-ILKOM)*, 3(1), 46–56. <https://doi.org/10.58602/jima-ilkom.v3i1.26>
- Lidya Musaffak, A., Maulida Hindrayani, K., & Idhom, M. (2025). Penerapan Metode Mean Shift Clustering Untuk Mengelompokkan Wilayah Berdasarkan Pengelolaan Sampah. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(3), 4790–4797. <https://doi.org/10.36040/jati.v9i3.13777>
- Ling, L. S., & Weiling, C. T. (2025). Enhancing Segmentation: A Comparative Study of Clustering Methods. *IEEE Access*, 13, 47418–47439. <https://doi.org/10.1109/ACCESS.2025.3550339>
- Lisnawati Sopiah, & Ria Haryatiningsih. (2023). Karakteristik Penduduk Miskin dan Penyebab Kemiskinan di Desa Sukagalih. *Jurnal Riset Ilmu Ekonomi Dan Bisnis*, 69–74. <https://doi.org/10.29313/jrieb.v3i1.1977>
- Maharana, K., Mondal, S., & Nemade, B. (2022). A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*, 3(1), 91–99. <https://doi.org/10.1016/j.gltp.2022.04.020>
- Punhani, A., Faujdar, N., Mishra, K. K., & Subramanian, M. (2022). Binning-Based Silhouette Approach to Find the Optimal Cluster Using K-Means. *IEEE Access*, 10, 115025–115032. <https://doi.org/10.1109/ACCESS.2022.3215568>
- Purba, W., Siawin, W., & . H. (2019). Implementasi Data Mining Untuk Pengelompokan Dan Prediksi Karyawan Yang Berpotensi Phk Dengan Algoritma K-Means Clustering. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima (JUSIKOM PRIMA)*, 2(2), 85–90. <https://doi.org/10.34012/jusikom.v2i2.429>
- Ren, Y., Pu, J., Yang, Z., Xu, J., Li, G., Pu, X., Yu, P. S., & He, L. (2025). Deep Clustering: A Comprehensive Survey. *IEEE Transactions on Neural Networks and Learning Systems*, 36(4), 5858–5878. <https://doi.org/10.1109/TNNLS.2024.3403155>
- Rianti, R., Andarsyah, R., & Awangga, R. M. (2024). Penerapan PCA dan Algoritma Clustering untuk Analisis Mutu Perguruan Tinggi di LLDIKTI Wilayah IV. *NUANSA INFORMATIKA*, 18(2), 67–77. <https://doi.org/10.25134/ilkom.v18i2.211>
- Rizuan, R., Haerani, E., Jasril, J., & Oktavia, L. (2023). Penerapan Algoritma Mean-Shift Pada Clustering Penerimaan Bantuan Pangan Non Tunai. *Journal of Computer System and Informatics (JoSYC)*, 4(4), 1019–1027. <https://doi.org/10.47065/josyc.v4i4.3876>
- Sangadah, S. K., Laut, L. T., & Jalunggono, G. (2020). Pengaruh faktor-faktor penyebab kemiskinan di Kabupaten Kebumen tahun 2009-2018. *DINAMIC: Directory Journal of Economic*, 2(1), 229–243.



- Sari, L. P., Nurjannah, N., & Yovita, I. (2022). Analysis Of The Effect Of Population, Human Development Index And Open Unemployment Rate On Poverty In Riau Province, 2002-2021. *Eqien - Jurnal Ekonomi Dan Bisnis*, 11(1), 567 – 574. <https://doi.org/10.34308/eqien.v11i1.721>
- Setiawan, Y. (2023). Data Mining berbasis Nearest Neighbor dan Seleksi Fitur untuk Deteksi Kanker Payudara. *Jurnal Informatika: Jurnal Pengembangan IT*, 8(2), 89–96. <https://doi.org/10.30591/jpit.v8i2.4994>
- Sinurat, R. P. P. (2023). Analisis Faktor-Faktor Penyebab Kemiskinan Sebagai Upaya Penanggulangan Kemiskinan Di Indonesia. *Jurnal Registratie*, 5(2), 87–103. <https://doi.org/10.33701/jurnalregistratie.v5i2.3554>
- Surbakti, N. K. (2021). Data Mining Pengelompokan Pasien Rawat Inap Peserta BPJS Menggunakan Metode Clustering (Studi Kasus : RSUD.Bangkalan). *Journal of Information and Technology*, 1(2), 47–53. <https://doi.org/10.32938/jitu.v1i2.1470>
- Tipwong, W., Sirikham, A., Konpang, J., Kirdpipat, P., Chongjarearn, Y., & Preutisrunyanont, O. (2025). Comparison of Unsupervised Learning Algorithms for K-Means Clustering Efficiency in IoE Sensor Detection Systems for Smart Agriculture. 2025 13th International Electrical Engineering Congress (IEECON), 1–5. <https://doi.org/10.1109/IEECON64081.2025.10987590>
- Utari, R. F., Insani, F., Agustian, S., & Afriyanti, L. (2024). Pengelompokan Data Pendistribusian Listrik Menggunakan Algoritma Mean Shift. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(3), 1015–1023. <https://doi.org/10.57152/malcom.v4i3.1428>
- Wang, J., Alroobaea, R., Baqasah, A. M., Althobaiti, A., & Kansal, L. (2023). Study on Library Management System Based on Data Mining and Clustering Algorithm. *Informatica*, 46(9). <https://doi.org/10.31449/inf.v46i9.3858>
- Zamri, N., Pairan, M. A., Azman, W. N. A. W., Abas, S. S., Abdullah, L., Naim, S., Tarmudi, Z., & Gao, M. (2022). A comparison of unsupervised and supervised machine learning algorithms to predict water pollutions. *Procedia Computer Science*, 204, 172–179. <https://doi.org/10.1016/j.procs.2022.08.021>
- Zhao, M., Jha, A., Liu, Q., Millis, B. A., Mahadevan-Jansen, A., Lu, L., Landman, B. A., Tyska, M. J., & Huo, Y. (2021). Faster Mean-shift: GPU-accelerated clustering for cosine embedding-based cell segmentation and tracking. *Medical Image Analysis*, 71, 102048. <https://doi.org/10.1016/j.media.2021.102048>