

**KLASTERISASI MASYARAKAT KURANG MAMPU DI KELURAHAN
DURIAN KOTA MEDAN UNTUK OPTIMALISASI PENYALURAN
BANTUAN SOSIAL MENGGUNAKAN ALGORITMA *ORDERING*
*POINTS TO IDENTIFY THE CLUSTERING STRUCTURE***

SKRIPSI

DISUSUN OLEH

SITI NURISMA SIREGAR

NPM. 2209020075



UMSU

Unggul | Cerdas | Terpercaya

**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

MEDAN

2026

**KLASTERISASI MASYARAKAT KURANG MAMPU DI KELURAHAN
DURIAN KOTA MEDAN UNTUK OPTIMALISASI PENYALURAN
BANTUAN SOSIAL MENGGUNAKAN ALGORITMA *ORDERING*
*POINTS TO IDENTIFY THE CLUSTERING STRUCTURE***

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
(S.Kom) dalam Program Studi Teknologi Informasi pada Fakultas Ilmu
Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara**

SITI NURISMA SIREGAR

NPM. 2209020075

**PROGRAM STUDI TEKNOLOGI INFROMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

MEDAN

2026

LEMBAR PENGESAHAN

Judul Skripsi : Klasterisasi Masyarakat Kurang Mampu di Kelurahan Durian Kota Medan untuk Optimalisasi Penyaluran Bantuan Sosial Menggunakan Algoritma Optics (*Ordering Points To Identify The Clustering Structure*)

Nama Mahasiswa : Siti Nurisma Siregar

NPM : 2209020075

Program Studi : Teknologi Informasi

Menyetujui
Komisi Pembimbing



(Yohanni Syahra, S.Si., M.Kom)
NIDN. 0129108201

Ketua Program Studi



(Fatma Sari Hutagalung, S.Kom, M.Kom)
NIDN. 0117019301



Dekan
(Dr. Al-Kho'warizmi, S.Kom., M.Kom.)
NIDN. 0127099201

PERNYATAAN ORISINALITAS

**Klasterisasi Masyarakat Kurang Mampu di Kelurahan Durian Kota Medan
untuk Optimalisasi Penyaluran Bantuan Sosial Menggunakan Algoritma
Optics (*Ordering Points To Identify The Clustering Structure*)**

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, 14 April 2026

Yang membuat pernyataan



Siti Nurisma Siregar

NPM. 2209020075

**PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN
AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama : Siti Nurisma Siregar
NPM : 22099020075
Program Studi : Teknologi Informasi
Karya Ilmiah : Skripsi

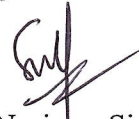
Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bedas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

**Klasterisasi Masyarakat Kurang Mampu di Kelurahan Durian Kota Medan
untuk Optimalisasi Penyaluran Bantuan Sosial Menggunakan Algoritma
Optics (*Ordering Points To Identify The Clustering Structure*)**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, 14 April 2026
Yang membuat pernyataan



Siti Nurisma Siregar
NPM. 2209020075

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap : Siti Nurisma Siregar
Tempat dan Tanggal Lahir : Medan, 24 September 2003
Alamat Rumah : Asrama Singgasana III K148, Medan Timur
Telepon/Faks/HP : 082273323954
E-mail : sitinurismasiregar@gmail.com
Instansi Tempat Kerja : -
Alamat Kantor : -

DATA PENDIDIKAN

SD : MIN Glugur Darat II TAMAT: 2016
SMP : MTS LAB UINSU TAMAT: 2019
SMA : SMKN 1 MEDAN TAMAT: 2022

KATA PENGANTAR



Pendahuluan

Penulis tentunya berterima kasih kepada berbagai pihak dalam dukungan serta doa dalam penyelesaian skripsi. Penulis juga mengucapkan terima kasih kepada:

1. Allah SWT, karena berkat limpahan rahmat dan karunia-Nya, penulis dapat menyelesaikan skripsi ini.
2. Bapak Prof. Dr. Akrim, M.Pd., Rektor Universitas Muhammadiyah Sumatera Utara (UMSU)
3. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
4. Ibu Dr. Firahmi Rizky, S.Kom., M.Kom. selaku Wakil Dekan I Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
5. Bapak Mhd. Basri, S.Si, M.Kom. selaku Wakil Dekan III Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
6. Ibu Fatma Sari Hutagalung S.Kom., M.Kom. selaku Ketua Program Studi Teknologi Informasi.
7. Bapak Okvi Nugroho, S.Kom., M.Kom selaku Sekretaris Program Studi Teknologi Informasi.
8. Ibu Yohanni Syahra, S.Si., M.Kom selaku Pembimbing Akademik penulis. Terima kasih telah menjadi sosok pembimbing yang tidak hanya membimbing secara akademik, tetapi juga memberikan perhatian tulus kepada penulis sebagai manusia yang sedang belajar dan bertumbuh.
9. Kedua Orang tua tersayang, Ayah dan Mama. Tidak ada kata yang cukup untuk menggambarkan betapa besar kasih sayang dan pengorbanan yang telah kalian berikan. Meskipun perjalanan ini tidak selalu mudah, kalian

tidak pernah sekalipun berhenti mendoakan, berjuang, dan memberikan yang terbaik untuk anak-anakmu. Tanpa gelar, tanpa pamrih hanya cinta yang tulus dan tak pernah padam. Setiap tetes keringat Ayah dan Mama adalah bahan bakar yang menggerakkan penulis untuk terus melangkah, bahkan di saat penulis hampir menyerah. Semoga skripsi ini menjadi salah satu bukti kecil bahwa pengorbanan kalian tidak pernah sia-sia. Semoga Ayah dan Mama selalu sehat, panjang umur, dan sempat menyaksikan sendiri setiap keberhasilan yang akan penulis raih karena semua itu, dari dulu hingga nanti, selalu untuk kalian.

10. Kepada Gaizka Pasya Dermawan, sahabat seperjuangan sejak tahun 2022. Terima kasih sudah menjadi seseorang yang tidak pernah mengeluh meski penulis berkali-kali merepotkan. Kamu selalu ada di saat penulis bingung, panik, bahkan hampir menyerah kamu hadir dengan ketenangan, waktu, tenaga, dan pikiranmu yang tidak pernah pelit untuk dibagikan. Mungkin kamu tidak tahu seberapa besar artimu dalam perjalanan ini, tapi penulis tahu dan tidak akan pernah lupa. Semoga semua kebaikan yang kamu berikan kembali kepadamu berlipat ganda, dan semoga setiap impian yang telah kita bicarakan bersama, satu per satu, benar-benar terwujud di kemudian hari.
11. Kepada Nadya Aulya Putri & Nabila Azura Putri, sahabat yang tidak hanya hadir di momen-momen menyenangkan, tetapi justru paling setia ada di saat-saat paling sulit. Terima kasih sudah menjadi tempat pulang ketika penulis tidak tahu harus bercerita kepada siapa. Kehadiran kalian bukan hal kecil, kalian adalah salah satu alasan penulis mampu bertahan dan terus melangkah sampai di titik ini. terima kasih telah menjadi bagian dari perjalanan ini dengan ketulusan yang tidak pernah penulis ragukan.
12. Kepada sahabat-sahabat SMK penulis, Faradila Sari Nasution, Dwi Resya Fadhilah, dan Allya. Kalian hadir jauh sebelum skripsi ini ada, dan tetap ada menemani setiap babak kehidupan yang terus berganti. terima kasih sudah tidak pernah benar-benar pergi, meskipun jarak dan kesibukan sering memisahkan. Canda tawa, cerita tengah malam, dan doa-doa tulus yang

kalian kirimkan diam-diam semuanya terasa, dan semuanya berarti lebih dari yang bisa penulis ungkapkan.

Karena kalian juga, penulis mampu sampai di titik ini. Semoga kita semua menemukan kebahagiaan yang sudah lama kita impikan, dan semoga jalan kita masih sering bertemu dalam keadaan yang jauh lebih baik dari hari ini.

13. Para rekan seperjuangan angkatan 2022 B1 TI. Bersama kalian adalah salah satu babak paling berharga dalam hidup penulis penuh malam panjang, tugas menumpuk, tekanan, sekaligus tawa yang tidak akan terlupakan. Kalian bukan sekadar teman sekelas, kalian adalah bagian dari perjalanan yang membentuk siapa penulis hari ini.

Meski setelah ini kita berjalan di jalan masing-masing, penulis berharap ikatan yang sudah kita bangun tidak ikut menghilang. Semoga kita semua menemukan kebahagiaan di setiap jalan yang kita pilih. Sampai jumpa di puncak kesuksesan kita masing-masing.

14. Terakhir tidak lupa, kepada diri saya sendiri. Terima kasih sudah memilih untuk bertahan dan berjuang untuk tetap ada hingga saat ini, serta menjadi perempuan yang kuat dan ikhlas atas segala perjalanan hidup tidak selalu mudah. Dengan adanya skripsi ini, telah berhasil membuktikan bahwa kamu bisa menyandang gelar S.Kom tepat waktu dan mampu membuktikan bahwa semua yang kamu lewati tidak sia-sia. Jadikan ini pengingat bahwa kamu selalu bisa, bahkan di hari-hari yang paling berat sekalipun. Apapun yang menanti setelah ini, jangan lupa menghargai dirimu. Kamu sudah sangat layak untuk bahagia.

15. Semua pihak yang terlibat langsung ataupun tidak langsung yang tidak dapat penulis ucapkan satu-persatu yang telah membantu penyelesaian skripsi ini.

**Klasterisasi Masyarakat Kurang Mampu di Kelurahan Durian Kota Medan
untuk Optimalisasi Penyaluran Bantuan Sosial Menggunakan Algoritma
Optics (*Ordering Points To Identify The Clustering Structure*)**

ABSTRAK

Penelitian ini bertujuan untuk mengelompokkan masyarakat kurang mampu di Kelurahan Durian Kota Medan guna mendukung optimalisasi penyaluran bantuan sosial menggunakan algoritma OPTICS (*Ordering Points to Identify the Clustering Structure*). Data yang digunakan merupakan data sosial-ekonomi masyarakat yang meliputi pendapatan, pengeluaran, pekerjaan, tingkat pendidikan, dan jumlah tanggungan dengan jumlah sebanyak 800 data, yang setelah preprocessing menjadi 789 data. Tahapan penelitian meliputi preprocessing data (*cleaning, encoding, feature engineering, dan normalisasi*), penentuan parameter menggunakan *k-distance plot* dan *grid search*, proses clustering dengan algoritma OPTICS, serta evaluasi menggunakan *Silhouette Index*. Hasil penelitian menunjukkan bahwa parameter optimal diperoleh pada nilai $\text{min_samples} = 15$, $\text{max_eps} = 0,3$, dan $\text{xi} = 0,030$ dengan nilai *Silhouette Index* sebesar 0,2409. Proses clustering menghasilkan 4 cluster utama yaitu tidak mampu, kurang mampu, mampu, dan sangat mampu, serta sejumlah data yang teridentifikasi sebagai noise. Algoritma OPTICS terbukti mampu mengidentifikasi struktur data dengan variasi kepadatan yang berbeda serta mendeteksi outlier secara otomatis. Hasil klasterisasi divisualisasikan menggunakan *reachability plot* dan *scatter plot* untuk mempermudah interpretasi. Penelitian ini diharapkan dapat memberikan kontribusi dalam meningkatkan ketepatan sasaran penyaluran bantuan sosial melalui pendekatan berbasis data.

Kata Kunci: *Clustering, OPTICS, Data Mining, Bantuan Sosial, Kemiskinan*

Clustering of Underprivileged Communities in Durian Village, Medan City for the Optimization of Social Assistance Distribution Using the OPTICS (Ordering Points to Identify the Clustering Structure) Algorithm

ABSTRACT

This study aims to cluster underprivileged communities in Durian Village, Medan City, to support the optimization of social assistance distribution using the OPTICS (Ordering Points to Identify the Clustering Structure) algorithm. The dataset used consists of socio-economic data, including income, expenditure, occupation, education level, and number of dependents, totaling 800 records, which were reduced to 789 records after preprocessing. The research stages include data preprocessing (cleaning, encoding, feature engineering, and normalization), parameter determination using k-distance plot and grid search, clustering using the OPTICS algorithm, and evaluation using the Silhouette Index. The results show that the optimal parameters were obtained with $\text{min_samples} = 15$, $\text{max_eps} = 0.3$, and $\text{xi} = 0.030$, resulting in a Silhouette Index value of 0.2409. The clustering process produced four main clusters: very poor, poor, moderate, and well-off groups, along with a number of data points identified as noise. The OPTICS algorithm proved effective in handling data with varying density levels and automatically detecting outliers. The clustering results were visualized using reachability plots and scatter plots to enhance interpretability. This study is expected to contribute to improving the accuracy of social assistance distribution through a data-driven approach.

Keywords: *Clustering, OPTICS, Data Mining, Social Assistance, Poverty*

DAFTAR ISI

LEMBAR PENGESAHAN	i
PERNYATAAN ORISINALITAS.....	ii
PERNYATAAN PERSETUJUAN PUBLIKASI.....	iii
RIWAYAT HIDUP	iv
KATA PENGANTAR.....	v
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian.....	5
BAB II LANDASAN TEORI	7
2.1 Masyarakat Kurang Mampu.....	7
2.2 Data Mining.....	7
2.3 Clustering	10
2.4 <i>OPTICS (Ordering Points to Identify the Clustering Structure)</i>	15
2.4.1 Keunggulan <i>OPTICS (Ordering Points to Identify the Clustering Structure)</i>	16
2.4.2 Tahapan <i>OPTICS (Ordering Points to Identify the Clustering Structure)</i> 16	
2.5 Kerangka Berpikir	20
2.6 Ringkasan Penelitian Terdahulu.....	23
2.7 Analisis GAP.....	25
BAB III METODOLOGI PENELITIAN	28
3.1 Jenis Penelitian	28
3.2 Teknik Pengumpulan Data	29
3.3 Arsitektur Penelitian.....	30
3.3.1 Pengumpulan Data.....	30
3.3.2 Preprocessing Data	31
3.3.3 Penentuan Kriteria dan Parameter	32

3.3.4	Proses dengan Algoritma <i>OPTICS</i>	33
3.3.5	Activity Diagram	36
3.3.6	Use Case Diagram	38
3.3.7	Analisis hasil dan perbandingan	40
BAB IV	HASIL DAN PEMBAHASAN	42
4.1.	Deskripsi data	42
4.2.	Preprocessing Data	43
4.2.1.	Data Cleaning	44
4.2.2.	Penghapusan Outlier.....	44
4.2.3.	Encoding Variabel Kategorikal	44
4.2.4.	Feature Engineering.....	45
4.2.5.	Normalisasi Data	46
4.3.	Penentuan Parameter Optics.....	48
4.3.1.	Analisis K-Distance Plot	48
4.3.2.	Grid Search Parameter.....	49
4.4.	Proses Algoritma OPTICS	50
4.4.1.	Perhitungan Core Distance dan Reachability Distance	50
4.4.2.	Reachability Plot	51
4.4.3.	Ekstraksi Cluster.....	51
4.5.	Perhitungan Manual OPTICS.....	52
4.5.1.	Data Sampel.....	52
4.5.2.	Hitung Jarak Euclidean Antar Titik.....	52
4.5.3.	Tentukan ϵ -Neighborhood dan Core Distance	53
4.5.4.	Hitung Reachability Distance	54
4.5.5.	Bangun Cluster Order.....	54
4.5.6.	Interpretasi Hasil.....	55
4.5.7.	Hasil Klasterisasi	56
4.6.	Analisis Karakteristik Setiap Cluster	58
4.6.1.	Cluster 0, Tidak Mampu (103 Masyarakat).....	58
4.6.2.	Cluster 1 (Kurang Mampu).....	58
4.6.3.	Cluster 2 (Mampu)	59
4.6.4.	Cluster 3 (Sangat Mampu).....	60
4.6.5.	Noise/Outlier (207 Masyarakat)	60
4.7.	Evaluasi Hasil Klasterisasi	61
4.7.1.	Silhouette Index	61
4.7.2.	Perbandingan dengan Penelitian Terdahulu	62

BAB V KESIMPULAN DAN SARAN	63
5.1 KESIMPULAN	63
5.2 SARAN	64
DAFTAR PUSTAKA	66

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	23
Tabel 4.1 Statistik Data Masyarakat Kelurahan Durian.....	42
Tabel 4.2 Distribusi Jenis Pekerjaan Masyarakat Kelurahan Durian.....	43
Tabel 4.3 Distribusi Tingkat Pendidikan Masyarakat Kelurahan Durian	43
Tabel 4.4 Encoding Tingkat Pendidikan.....	44
Tabel 4.5 Encoding Jenis Pekerjaan.....	45
Tabel 4.6 Ringkasan Hasil Preprocessing Data	46
Tabel 4.7 Hasil Grid Search Parameter OPTICS (Kombinasi Terbaik)	50
Tabel 4.8 Data Sampel untuk Perhitungan Manual OPTICS.....	52
Tabel 4.9 Hasil Cluster Order dari perhitungan manual OPTICS	55
Tabel 4.10 Ringkasan Hasil Klasterisasi OPTICS	56
Tabel 4.11 Statistik Cluster 0 (Tidak Mampu)	58
Tabel 4.12 Statistik Cluster 1 (Kurang Mampu).....	59
Tabel 4.13 Statistik Cluster 2 (Mampu).....	59
Tabel 4.14 Statistik Cluster 3 (Sangat Mampu).....	60
Tabel 4.15 Hasil Evaluasi Silhouette Index	61

DAFTAR GAMBAR

Gambar 2.1 Tahapan data mining menggunakan CRISP-DM.....	8
Gambar 2. 2 Data Masyarakat Warga Kelurahan Durian	29
Gambar 3.1 Arsitektur Penelitian.....	30
Gambar 3.2 Data masyarakat kurang mampu	31
Gambar 3.3 Data masyarakat kurang mampu lanjut.....	31
Gambar 3. 4 Flowchart OPTICS	34
Gambar 3.5 Activity Diagram.....	36
Gambar 3.6 Use Case Diagram.....	38
Gambar 4.1 Hasil Pratinjau Pra-Proses pada Aplikasi.....	47
Gambar 4.2 Distribusi Variabel Sebelum dan Sesudah Preprocessing.....	47
Gambar 4.3 Preview 10 Data Pertama Setelah Preprocessing.....	48
Gambar 4.4 K - Distance Plot untuk Penentuan Nilai Optimal ϵ	49
Gambar 4.5 Hasil Auto Tune Parameter dan Tabel Grid Search OPTICS	49
Gambar 4. 6 Reachability Plot Hasil OPTICS (min_samples=15, max_eps=0,3)	51
Gambar 4.7 Hasil Klasterisasi OPTICS, Profil Tiap Cluster	56
Gambar 4.8 Scatter Plot Hasil Klasterisasi OPTICS (Pendapatan vs Pengeluaran)	57
Gambar 4. 9 Distribusi Jumlah Masyarakat Durian per Cluster	57

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Kemiskinan merupakan permasalahan kompleks yang dihadapi oleh berbagai negara di dunia, termasuk Indonesia. Menurut data Badan Pusat Statistik (BPS) tahun 2024, tingkat kemiskinan di Indonesia masih menjadi tantangan besar dalam pembangunan nasional, terutama di wilayah perkotaan yang mengalami urbanisasi tinggi (Dwitra Gusti Alriscki & Fauzan, 2024). Kota Medan sebagai salah satu kota metropolitan di Indonesia memiliki dinamika kemiskinan yang kompleks, dengan konsentrasi masyarakat kurang mampu yang tersebar di berbagai kelurahan. Kelurahan Durian merupakan salah satu wilayah di Kota Medan yang memiliki populasi masyarakat kurang mampu cukup signifikan, sehingga memerlukan perhatian khusus dalam pengelolaan program bantuan sosial.

Pemerintah Indonesia telah mengalokasikan anggaran bantuan sosial yang cukup besar untuk mengatasi kemiskinan melalui berbagai program seperti Program Keluarga Harapan (PKH), Bantuan Pangan Non Tunai (BPNT), dan Program Indonesia Pintar (Jhos Franklin Kemit, 2024). Efektivitas penyaluran bantuan masih menjadi persoalan yang perlu diperbaiki (Ferdiansah & Kriswibowo, 2023). Permasalahan yang sering terjadi adalah ketidaktepatan sasaran penerima bantuan, distribusi yang tidak merata, dan kurangnya klasifikasi yang jelas mengenai tingkat kebutuhan masyarakat. Data masyarakat kurang mampu yang dimiliki oleh pemerintah daerah seringkali bersifat heterogen dan kompleks, mencakup berbagai aspek sosial-ekonomi seperti pendapatan keluarga, pengeluaran bulanan, jumlah tanggungan, pekerjaan, tingkat pendidikan.

Data mining merupakan teknologi yang dapat digunakan untuk mengekstraksi informasi dan pengetahuan tersembunyi dari kumpulan data yang besar dan kompleks (Darmawan et al., 2022). Salah satu teknik dalam data mining yang dapat diterapkan untuk mengatasi permasalahan penyaluran bantuan sosial adalah klasterisasi atau clustering. Klasterisasi memungkinkan pengelompokan masyarakat kurang mampu berdasarkan kemiripan karakteristik sosial-ekonomi sehingga dapat dikategorikan menjadi kelompok masyarakat mampu, kurang mampu, dan tidak mampu. Dengan mengklasterisasi masyarakat kurang mampu di Kelurahan Durian, pihak kelurahan dan pemerintah daerah dapat menyusun strategi penyaluran bantuan yang lebih terarah, mengalokasikan sumber daya bantuan sosial secara optimal, serta merancang program pemberdayaan yang sesuai dengan karakteristik dan kebutuhan masing-masing kelompok masyarakat.

Berbagai algoritma klasterisasi telah dikembangkan untuk mengelompokkan data, namun tidak semua algoritma cocok untuk menangani data masyarakat kurang mampu yang memiliki karakteristik beragam dan kepadatan yang bervariasi. Algoritma klasterisasi berbasis partisi seperti K-Means memerlukan penentuan jumlah klaster di awal dan sensitif terhadap outlier. Sementara algoritma berbasis hierarki cenderung memiliki kompleksitas komputasi yang tinggi untuk dataset besar. Oleh karena itu, diperlukan algoritma klasterisasi yang mampu menangani data dengan variasi kepadatan berbeda, tidak memerlukan penentuan jumlah klaster di awal, dan dapat mengidentifikasi outlier secara otomatis.

Algoritma OPTICS (Ordering Points To Identify the Clustering Structure) merupakan metode klasterisasi density-based yang mampu mengidentifikasi

kelompok data dengan variasi kepadatan berbeda tanpa menentukan jumlah kluster di awal. OPTICS bekerja dengan mengurutkan objek berdasarkan *core distance* dan *reachability distance*, sehingga dapat mendeteksi kelompok masyarakat dengan tingkat kemiskinan yang bervariasi serta mengidentifikasi *outlier* secara otomatis. Keunggulan OPTICS dibandingkan dengan algoritma DBSCAN adalah kemampuannya dalam menangani kluster dengan kepadatan yang bervariasi dalam satu dataset, memberikan fleksibilitas dalam ekstraksi kluster dengan berbagai tingkat kepadatan, dan menghasilkan struktur kluster yang lebih informatif.

Dalam penelitian ini, algoritma OPTICS akan diterapkan untuk mengklusterisasi masyarakat kurang mampu di Kelurahan Durian Kota Medan. Data yang digunakan terdiri dari pendapatan keluarga, pengeluaran bulanan, pekerjaan, jumlah tanggungan, tingkat pendidikan. Hasil klusterisasi diharapkan dapat memberikan pengelompokan masyarakat berdasarkan tingkat kemiskinan sebagai acuan bagi Pemerintah Kota Medan dan Dinas Sosial Kota Medan dalam merumuskan strategi penyaluran bantuan sosial yang tepat sasaran. Dengan demikian, penelitian ini diharapkan dapat berkontribusi dalam meningkatkan efektivitas dan efisiensi program bantuan sosial melalui pendekatan berbasis data dan teknologi informasi.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian saya adalah:

1. Bagaimana mengelompokkan masyarakat kurang mampu di Kelurahan Durian Kota Medan ke dalam kluster berdasarkan karakteristik sosial-ekonomi menggunakan algoritma OPTICS?

2. Bagaimana kinerja algoritma OPTICS dalam mengklusterisasi data masyarakat kurang mampu berdasarkan metrik evaluasi Silhouette Index?
3. Bagaimana menentukan nilai optimal ϵ dan MinPts pada algoritma OPTICS sehingga menghasilkan Silhouette Index tertinggi untuk data kemiskinan di Kelurahan Durian?

1.3 Batasan Masalah

Penelitian ini memiliki batasan permasalahan yang dibahas agar dapat menyelesaikan permasalahan utama. Berikut ini batasan masalah penelitian saya:

1. Lokasi penelitian terbatas pada masyarakat kurang mampu di Kelurahan Durian Kota Medan.
2. Data yang digunakan dalam penelitian ini adalah data masyarakat kurang mampu yang diperoleh dari Kelurahan Durian Kota Medan yang mencakup variabel pendapatan keluarga, pengeluaran bulanan, pekerjaan, jumlah tanggungan, dan tingkat pendidikan.
3. Algoritma klusterisasi yang digunakan adalah OPTICS (*Ordering Points to Identify the Clustering Structure*)
4. Metrik evaluasi menggunakan metrik Silhouette Index.
5. Implementasi algoritma dilakukan menggunakan bahasa pemrograman Python.
6. Penelitian ini tidak mencakup implementasi sistem penyaluran bantuan sosial secara real-time, namun fokus pada proses klusterisasi dan visualisasi hasil untuk mendukung pengambilan keputusan.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah ditetapkan, maka tujuan penelitian saya adalah:

1. Mengelompokkan masyarakat kurang mampu di Kelurahan Durian Kota Medan ke dalam kluster berdasarkan karakteristik sosial-ekonomi menggunakan algoritma OPTICS.
2. Mengevaluasi kinerja algoritma OPTICS dalam mengklusterisasi data masyarakat kurang mampu menggunakan metrik evaluasi clustering (Silhouette Index).
3. Untuk menentukan nilai ϵ dan MinPts yang optimal pada OPTICS dengan menganalisis grafik k-distance plot serta menguji beberapa kombinasi parameter, sehingga dapat memperoleh Silhouette Index tertinggi sehingga klusterisasi paling akurat dan dapat diandalkan untuk mengidentifikasi kelompok kemiskinan di Kelurahan Durian.

1.5 Manfaat Penelitian

Adapun manfaat penelitian yang dihasilkan dari penelitian ini adalah sebagai berikut:

1. Dapat Memberikan kontribusi ilmiah dalam penerapan algoritma OPTICS untuk klusterisasi data masyarakat kurang mampu dan memperkaya literatur akademik tentang *density-based clustering* pada data sosial-ekonomi.
2. Dapat Meningkatkan akurasi dan keadilan distribusi bantuan sosial agar tepat sasaran melalui pengelompokan kategori sangat miskin, miskin sedang, dan hampir miskin.

3. Dapat menjadi bahan rekomendasi bagi pihak kelurahan durian dalam menentukan wilayah masyarakat kurang mampu untuk optimalisasi bantuan sosial.

BAB II

LANDASAN TEORI

2.1 Masyarakat Kurang Mampu

Masyarakat kurang mampu atau masyarakat miskin merupakan kelompok masyarakat yang memiliki keterbatasan dalam memenuhi kebutuhan dasar hidupnya seperti pangan, sandang, papan, kesehatan, dan pendidikan. Menurut Badan Pusat Statistik (BPS), kemiskinan adalah kondisi seseorang atau sekelompok orang yang tidak mampu memenuhi kebutuhan dasarnya untuk mempertahankan dan mengembangkan kehidupan yang bermartabat (BPS, 2024).

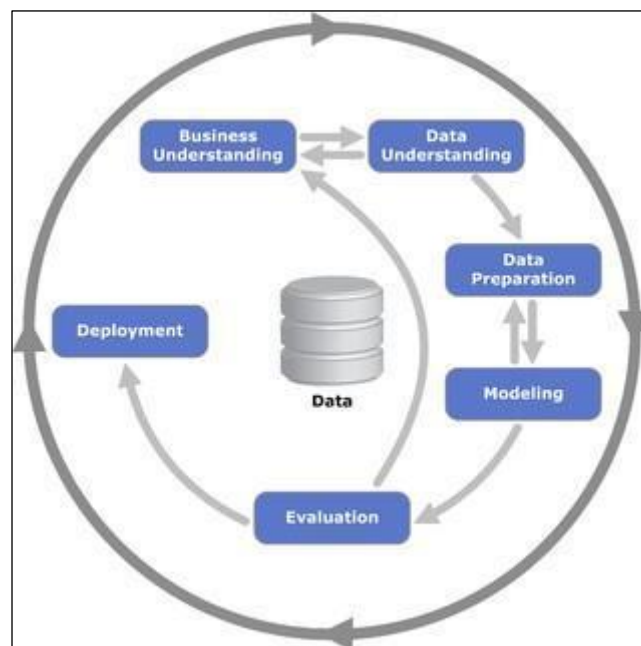
Karakteristik masyarakat kurang mampu dapat diidentifikasi melalui berbagai indikator sosial-ekonomi yang meliputi tingkat pendapatan keluarga yang rendah, pengeluaran konsumsi yang terbatas, tingkat pendidikan yang rendah, dan jumlah tanggungan keluarga yang banyak. Identifikasi dan klasifikasi masyarakat kurang mampu menjadi kunci keberhasilan program bantuan sosial, di mana penggunaan data dan teknologi informasi dalam proses identifikasi dapat meningkatkan akurasi penentuan penerima bantuan dan mengurangi kesalahan dalam pendataan (Dwitra Gusti Alriski & Fauzan, 2024).

2.2 Data Mining

Data mining merupakan proses ekstraksi informasi dan pengetahuan yang berguna dari kumpulan data yang besar dengan menggunakan teknik komputasi dan statistik (Darmawan et al., 2022). Menurut penelitian darmawan yang berjudul “Penerapan Data Mining Menggunakan Algoritma Apriori Untuk Menentukan Pola Golongan Penyandang Masalah Kesejahteraan Sosial” data mining adalah proses

menemukan pola yang menarik dan pengetahuan dari sejumlah besar data yang tersimpan dalam basis data, data warehouse, atau media penyimpanan informasi lainnya. Data mining juga dapat didefinisikan sebagai proses otomatis atau semi-otomatis untuk menemukan pola, asosiasi, perubahan, anomali, dan struktur yang signifikan secara statistik dalam data.

Proses data mining melibatkan beberapa tahapan yang sistematis untuk menghasilkan pengetahuan yang berguna dari data (Syahra et al., 2025). Tahapan-tahapan tersebut umumnya mengikuti metodologi *Cross-Industry Standard Process for Data Mining (CRISP-DM)* yang terdiri dari enam fase utama yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment* (Schröer et al., 2021). CRISP-DM merupakan standar *de-facto* yang paling banyak digunakan dalam proyek data mining karena sifatnya yang iteratif, fleksibel, dan independen terhadap industri maupun teknologi yang digunakan:



Gambar 2.1 Tahapan data mining menggunakan CRISP-DM

Sumber: (Abdusyukur, 2023)

1. *Business Understanding* (Pemahaman Bisnis)

Tahap pertama fokus pada pemahaman tujuan dan kebutuhan proyek dari perspektif bisnis atau domain aplikasi. Aktivitas utama meliputi penentuan tujuan bisnis, penilaian situasi dan sumber daya yang tersedia, penentuan tujuan data mining, serta perumusan rencana proyek. Dalam konteks penelitian ini, fase *business understanding* mencakup identifikasi permasalahan ketidaktepatan sasaran penyaluran bantuan sosial dan penentuan tujuan clustering untuk mengelompokkan masyarakat kurang mampu.

2. *Data Understanding* (Pemahaman Data)

Tahap kedua melibatkan pengumpulan data awal dan eksplorasi untuk memahami karakteristik data. Aktivitas meliputi koleksi data, deskripsi data, eksplorasi data melalui analisis statistik deskriptif dan visualisasi, serta verifikasi kualitas data untuk mengidentifikasi masalah seperti missing values atau inkonsistensi.

3. *Data Preparation* (Persiapan Data)

Tahap ketiga merupakan tahap menyiapkan dataset final yang akan digunakan untuk pemodelan. Tahap ini mencakup seleksi data, data cleaning, konstruksi atribut baru jika diperlukan, integrasi data dari berbagai sumber, dan formatting data ke dalam format yang sesuai. Fase ini sering menjadi tahap yang paling memakan waktu dalam proyek data mining.

4. *Modeling* (Pemodelan)

Tahap keempat adalah pemilihan dan penerapan teknik pemodelan yang sesuai. Aktivitas meliputi pemilihan algoritma, pembangunan model, dan penyesuaian

parameter untuk mengoptimalkan kinerja model. Dalam penelitian ini, algoritma OPTICS dipilih sebagai teknik clustering untuk mengelompokkan masyarakat kurang mampu.

5. *Evaluation* (Evaluasi)

Tahap kelima adalah evaluasi model yang telah dibangun untuk memastikan model memenuhi tujuan bisnis. Evaluasi dilakukan menggunakan metrik yang sesuai dengan jenis teknik data mining yang digunakan, serta review terhadap proses untuk menentukan langkah selanjutnya.

6. *Deployment* (Implementasi)

Tahap terakhir adalah implementasi hasil data mining ke dalam lingkungan operasional. Tahap ini meliputi perencanaan deployment, monitoring dan maintenance model, pembuatan laporan akhir, dan review proyek secara keseluruhan. Tujuan dari fase ini adalah memastikan pengetahuan yang diperoleh dapat digunakan secara efektif untuk mendukung pengambilan keputusan.

Sifat iteratif *CRISP-DM* memungkinkan peneliti untuk kembali ke fase sebelumnya jika diperlukan, menjadikannya metodologi yang fleksibel dan adaptif terhadap kebutuhan proyek yang dinamis (Abdusyukur, 2023).

2.3 Clustering

Clustering atau klasterisasi adalah teknik data mining untuk mengelompokkan data ke dalam kelompok-kelompok (cluster) berdasarkan kemiripan karakteristik tanpa adanya label kelas yang telah ditentukan sebelumnya (Adiguno et al., 2022). Data clustering adalah proses membagi dataset menjadi kelompok atau cluster, di mana instance dalam cluster yang sama memiliki

kemiripan satu sama lain, sedangkan instance dari cluster yang berbeda memiliki perbedaan (Sugianto & Bokings, 2021). Tujuan dari data clustering adalah untuk menemukan pola tersembunyi, struktur, atau hubungan dalam data yang dapat membantu dalam meringkas dataset yang besar dan kompleks serta memberikan representasi tingkat tinggi yang lebih mudah diinterpretasi dan dianalisis (Bahtiar et al., 2023).

Cluster yang baik memiliki karakteristik yaitu homogenitas internal yang tinggi (objek dalam satu cluster sangat mirip satu sama lain) dan heterogenitas eksternal yang tinggi (objek antar cluster sangat berbeda). Prinsip dasar clustering adalah memaksimalkan intra-cluster similarity (kemiripan dalam cluster) dan meminimalkan inter-cluster similarity (kemiripan antar cluster) (K-means et al., 2023). Clustering termasuk dalam kategori unsupervised learning karena tidak memerlukan data training dengan label kelas yang sudah diketahui, melainkan leverage sejumlah kecil data berlabel dalam kombinasi dengan set data tak berlabel yang lebih besar.

Algoritma clustering dapat dikategorikan berdasarkan metodologi yang digunakan. Berikut adalah kategori-kategori utama dari metode clustering (Jureksi et al., 2024) (Abdusyukur, 2023):

1. *Partitioning Methods* (Metode Partisi)

Metode partisi seperti K-Means, K-Medoids, dan PAM (*Partitioning Around Medoids*) bertujuan untuk membagi dataset menjadi cluster-cluster yang berbeda dengan cara mempartisi data ke dalam sejumlah kelompok yang telah ditentukan sebelumnya (Dwitra Gusti Alriscki & Fauzan, 2024). K-Means merupakan algoritma partisi yang paling populer dan banyak digunakan dalam

berbagai aplikasi data mining karena kesederhanaan konsepnya, efisiensi komputasinya dalam menangani dataset berskala besar, serta kemampuannya untuk menghasilkan hasil clustering dalam waktu yang relatif cepat. Algoritma K-Means bekerja dengan cara iteratif untuk meminimalkan jarak antara setiap data point dengan centroid cluster terdekat, sehingga menghasilkan cluster yang kompak dan terpisah dengan baik. Namun demikian, K-Means memiliki beberapa keterbatasan yang perlu diperhatikan, yaitu sensitivitas terhadap pemilihan posisi centroid awal yang dapat mempengaruhi hasil akhir clustering, ketergantungan pada penentuan jumlah cluster k yang harus ditentukan secara manual di awal proses, serta asumsi bahwa cluster memiliki bentuk spherical dengan ukuran yang relatif sama. Kelemahan utama dari metode partisi secara umum adalah ketidakmampuannya dalam menangani cluster dengan bentuk yang tidak beraturan atau *non-spherical*, sensitivitas yang tinggi terhadap keberadaan outlier atau noise dalam data yang dapat mengganggu proses clustering, serta kecenderungan untuk menghasilkan cluster dengan ukuran yang seimbang meskipun distribusi data yang sebenarnya tidak demikian (Fitriyah et al., 2023).

2. *Hierarchical Methods* (Metode Hierarki)

Hierarchical clustering mengorganisir data menjadi representasi berbasis pohon, tanpa memerlukan spesifikasi jumlah cluster di awal. Terdapat dua pendekatan utama yaitu agglomerative (*bottom-up*) yang dimulai dengan setiap objek sebagai cluster tunggal kemudian digabungkan secara bertahap, dan divisive (*top-down*) yang dimulai dengan satu cluster besar kemudian dipecah secara bertahap. Keuntungan metode hierarki adalah tidak perlu menentukan

jumlah cluster di awal dan menghasilkan visualisasi dendogram yang informatif, namun memiliki kompleksitas komputasi yang tinggi untuk dataset besar (Tugas Setiyawan & Shouni Barkah, 2025).

3. *Density-Based Methods* (Metode Berbasis Kepadatan)

Metode berbasis kepadatan mengelompokkan data berdasarkan area dengan kepadatan tinggi yang dipisahkan oleh area dengan kepadatan rendah. Algoritma yang termasuk dalam kategori ini adalah DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*), OPTICS (*Ordering Points To Identify the Clustering Structure*), dan DENCLUE. Keunggulan metode ini adalah kemampuan menemukan cluster dengan bentuk arbitrary, tidak memerlukan spesifikasi jumlah cluster di awal, dan dapat mengidentifikasi outlier atau noise secara otomatis. Namun, metode ini memiliki kesulitan dalam menangani data dengan variasi kepadatan yang signifikan dan sensitif terhadap pemilihan parameter (Klaster Berbasis Kepadatan Dengan Dbscan Dan Optics & Salman, 2023).

4. *Grid-Based Methods* (Metode Berbasis Grid)

Metode berbasis grid membagi ruang data menjadi struktur grid yang terbatas jumlahnya. Algoritma yang termasuk dalam kategori ini adalah STING (*Statistical Information Grid*), CLIQUE (*Clustering In QUEst*), dan *WaveCluster*. Keuntungan utama metode ini adalah kecepatan pemrosesan yang tinggi karena operasi clustering dilakukan pada grid bukan pada seluruh data point, namun kualitas hasil clustering sangat bergantung pada ukuran grid yang dipilih.

5. *Model-Based Methods* (Metode Berbasis Model)

Metode berbasis model mengasumsikan bahwa data dihasilkan dari campuran beberapa distribusi probabilitas yang mendasari struktur data tersebut. Pendekatan ini didasarkan pada teori statistik yang kuat dan memberikan kerangka kerja probabilistik untuk proses clustering. Algoritma yang termasuk dalam kategori ini adalah *Expectation-Maximization* (EM) yang merupakan metode iteratif untuk menemukan parameter distribusi yang optimal, *Gaussian Mixture Model* (GMM) yang memodelkan data sebagai campuran dari beberapa distribusi Gaussian, dan Bayesian Clustering yang menggunakan pendekatan Bayesian untuk inferensi probabilitas keanggotaan cluster. Metode berbasis model memiliki keunggulan dalam kemampuannya menangani cluster dengan bentuk ellipsoidal yang lebih fleksibel dibandingkan bentuk spherical, memberikan probabilitas keanggotaan (*soft clustering*) untuk setiap data point sehingga memungkinkan analisis ketidakpastian dalam pengelompokan, serta menyediakan kerangka kerja statistik yang kuat untuk interpretasi hasil clustering. Namun demikian, metode ini memiliki keterbatasan yaitu memerlukan asumsi tentang bentuk distribusi data yang seringkali sulit dipenuhi dalam praktik pada data real-world, kompleksitas komputasi yang tinggi terutama untuk dataset besar dengan banyak dimensi, sensitivitas terhadap inisialisasi parameter awal yang dapat mempengaruhi konvergensi algoritma, serta kesulitan dalam menentukan jumlah komponen distribusi yang optimal untuk merepresentasikan struktur data dengan baik (Abdusyukur, 2023).

2.4 *OPTICS (Ordering Points to Identify the Clustering Structure)*

OPTICS (Ordering Points to Identify the Clustering Structure) adalah algoritma clustering berbasis kepadatan yang dikembangkan oleh Ankerst, Breunig, Kriegel, dan Sander pada tahun 1999 sebagai perbaikan dari algoritma DBSCAN (Abdul et al., 2021). Algoritma ini tidak menghasilkan clustering dari dataset secara eksplisit, tetapi membuat *augmented ordering* dari database yang merepresentasikan struktur clustering berbasis kepadatan (Tugas Setiyawan & Shouni Barkah, 2025). *Cluster-ordering* ini berisi informasi yang setara dengan density-based clusterings yang sesuai dengan berbagai rentang pengaturan parameter, menjadikannya basis yang serbaguna untuk analisis cluster baik otomatis maupun interaktif.

Tidak seperti DBSCAN yang menghasilkan satu set cluster berdasarkan parameter epsilon dan MinPts yang tetap, OPTICS menghasilkan urutan ordering dari data points yang merepresentasikan struktur clustering berbasis kepadatan untuk berbagai nilai parameter (Hastuti et al., 2024). Konsep dasar OPTICS dibangun dari dua metrik utama yaitu core distance dan reachability distance. Core distance dari suatu point p adalah jarak minimum yang diperlukan agar p menjadi core point, yaitu memiliki minimal MinPts objek dalam neighborhood-nya (Tugas Setiyawan & Shouni Barkah, 2025). Reachability distance dari point q terhadap point p adalah maksimum antara core distance p dan jarak euclidean antara p dan q , yang merepresentasikan jarak terkecil yang harus ditempuh untuk mencapai q dari p melalui neighborhood yang padat (Rabbani et al., 2025).

2.4.1 Keunggulan OPTICS (Ordering Points to Identify the Clustering Structure)

Keunggulan OPTICS dibandingkan dengan algoritma clustering lainnya terletak pada beberapa aspek (Klaster Berbasis Kepadatan Dengan Dbscan Dan Optics & Salman, 2023). Pertama, OPTICS dapat menemukan cluster dengan kepadatan yang bervariasi dalam satu dataset tanpa perlu menjalankan algoritma berkali-kali dengan parameter yang berbeda. Kedua, OPTICS tidak secara langsung menghasilkan cluster, tetapi menghasilkan ordering augmented dengan informasi reachability distance yang dapat digunakan untuk mengekstrak clustering dengan berbagai parameter. Ketiga, reachability plot yang dihasilkan OPTICS memberikan visualisasi yang intuitif tentang struktur clustering dalam data, di mana lembah dalam plot merepresentasikan cluster dan ketinggian lembah mengindikasikan kepadatan cluster. Keempat, OPTICS dapat mendeteksi dan mengidentifikasi outlier atau noise secara otomatis (Fadilah & Wijayanto, 2023).

2.4.2 Tahapan OPTICS (Ordering Points to Identify the Clustering Structure)

Algoritma OPTICS bekerja melalui beberapa tahapan sistematis untuk menghasilkan cluster-order dari dataset (Tugas Setiyawan & Shouni Barkah, 2025). Berikut adalah tahapan detail dari algoritma OPTICS:

1. Inisialisasi Parameter

Tahap pertama adalah menentukan dua parameter utama yang dibutuhkan oleh OPTICS (Fadilah & Wijayanto, 2023). Parameter pertama adalah epsilon (ϵ) yang merupakan jarak maksimum untuk mendefinisikan neighborhood dari suatu point (Klaster Berbasis Kepadatan Dengan Dbscan Dan Optics & Salman, 2023). Parameter kedua adalah MinPts yang merupakan jumlah minimum

points yang harus ada dalam epsilon-neighborhood agar suatu point dapat dianggap sebagai core point (Hastuti et al., 2024). Pemilihan parameter ini mempengaruhi hasil clustering, namun OPTICS lebih robust terhadap pemilihan epsilon dibandingkan dengan DBSCAN karena menghasilkan ordering untuk berbagai nilai parameter.

2. Perhitungan Core Distance

Untuk setiap point p dalam dataset, core distance dihitung sebagai jarak dari point p ke neighbor terdekat yang ke- $MinPts$ ($MinPts$ -th nearest neighbor). Perhitungan ini dilakukan dengan cara mengurutkan semua neighbor dalam epsilon-neighborhood berdasarkan jaraknya terhadap point p , kemudian mengambil nilai jarak pada urutan ke- $MinPts$ sebagai core distance. Jika jumlah points dalam epsilon-neighborhood dari p kurang dari $MinPts$, maka point tersebut tidak memenuhi syarat sebagai core point dan core distance-nya bernilai undefined atau tak terdefinisi. Core distance merepresentasikan tingkat kepadatan lokal di sekitar point p , di mana nilai core distance yang kecil mengindikasikan bahwa point p berada di area yang padat dengan banyak neighbor di sekitarnya, sedangkan nilai core distance yang besar mengindikasikan bahwa point p berada di area yang jarang atau sparse dengan sedikit neighbor di sekitarnya. Core distance dari suatu titik p adalah jarak minimum yang dibutuhkan agar titik p memiliki setidaknya $MinPts$ tetangga di dalam radius tersebut. Core distance didefinisikan sebagai:

$$\text{core-dist}_{\epsilon, MinPts}(p) = \begin{cases} \text{UNDEFINED} & \text{jika } |N_{\epsilon}(p)| < MinPts \\ d_{MinPts}(p) & \text{jika } |N_{\epsilon}(p)| \geq MinPts \end{cases} \dots \dots \dots (2.1)$$

Di mana:

- a. $N_{\epsilon}(p)$ adalah himpunan titik-titik yang berada dalam radius ϵ dari titik p

- b. $|N_\epsilon(p)|$ adalah jumlah titik dalam lingkungan ϵ dari p
- c. $d_{\text{MinPts}}(p)$ adalah jarak dari p ke tetangga terdekat ke-MinPts (jarak terkecil yang memenuhi syarat MinPts)

Jika jumlah tetangga dalam radius ϵ kurang dari MinPts, maka titik p bukan merupakan core point dan core distance-nya bernilai UNDEFINED.

3. Perhitungan Reachability Distance

Reachability distance (rd) dari point q terhadap point p dihitung sebagai nilai maksimum antara core distance dari p dan jarak Euclidean antara p dan q .

Secara matematis, dinyatakan sebagai:

$$\text{rd}(q, p) = \max(\text{core distance}(p), \text{distance}(p, q)) \dots \dots \dots (2.2)$$

Di mana:

- a. $d(p, o)$ adalah jarak Euclidean antara titik p dan o
- b. $\text{core-dist}(o)$ adalah *core distance* dari titik o

Reachability distance merepresentasikan jarak minimum yang diperlukan untuk "mencapai" point q dari p melalui neighborhood yang padat. Konsep ini penting karena mempertimbangkan kepadatan lokal dalam pengukuran jarak antar points.

4. Pembangunan Cluster-Order

OPTICS membangun cluster-order dengan memproses setiap point dalam dataset secara iteratif. Algoritma dimulai dengan memilih point yang belum diproses, menghitung core distance dan reachability distance untuk point tersebut dan neighbors-nya, kemudian menyimpan informasi ini dalam struktur data ordered file. Point-point diurutkan berdasarkan nilai reachability distance dalam priority queue, dan point dengan reachability distance terkecil diproses

terlebih dahulu. Proses ini diulang hingga semua points telah diproses, menghasilkan ordering linear dari seluruh dataset yang merepresentasikan struktur clustering berbasis kepadatan.

5. Konstruksi Reachability Plot

Setelah cluster-order dibangun, OPTICS menghasilkan reachability plot untuk visualisasi. Reachability plot adalah grafik yang menampilkan urutan points pada sumbu x dan reachability distance pada sumbu y. Lembah dalam reachability plot merepresentasikan cluster, di mana kedalaman lembah mengindikasikan kepadatan cluster (semakin dalam lembah, semakin padat cluster), dan lebar lembah mengindikasikan jumlah points dalam cluster. Visualisasi ini memberikan insight yang kaya tentang struktur data dan memungkinkan identifikasi cluster dengan berbagai tingkat kepadatan.

6. Ekstraksi Cluster

Tahap terakhir adalah mengekstrak cluster dari cluster-order yang telah dibangun. Terdapat beberapa metode untuk ekstraksi cluster, termasuk metode threshold-based yang menggunakan nilai threshold pada reachability distance untuk memisahkan cluster. Alternatif lain adalah menggunakan algoritma ekstraksi otomatis seperti OPTICS-Xi yang dapat mengidentifikasi cluster secara otomatis berdasarkan steepness dalam reachability plot. Hasil akhir adalah assignment dari setiap point ke cluster tertentu atau label sebagai noise jika point tersebut tidak termasuk dalam cluster manapun.

Dari reachability plot yang telah dibangun, kluster maupun anomali dapat diekstraksi. Titik yang memiliki nilai reachability distance sangat tinggi atau bernilai UNDEFINED dianggap sebagai noise atau anomali, karena titik

tersebut tidak dapat dijangkau dari titik core mana pun dalam radius ε . Secara formal, suatu titik p dikategorikan sebagai anomali apabila:

$$\text{reach-dist}(p, o) > \varepsilon \text{ untuk semua } o \in N_\varepsilon \dots \dots \dots (2.3)$$

atau ketika core distance-nya bernilai UNDEFINED karena $|N_\varepsilon(p)| < \text{MinPts}$.

2.5 Kerangka Berpikir

Kemiskinan merupakan permasalahan kompleks yang dihadapi oleh berbagai wilayah di Indonesia, termasuk Kelurahan Durian Kota Medan (Mayasari & Nugraha, 2023). Tingkat kemiskinan yang masih tinggi dan penyaluran bantuan sosial yang belum optimal menjadi tantangan besar dalam upaya pengentasan kemiskinan. Permasalahan utama yang dihadapi adalah ketidaktepatan sasaran penyaluran bantuan sosial, distribusi yang tidak merata, serta kurangnya klasifikasi yang jelas mengenai tingkat kebutuhan masyarakat kurang mampu. Data masyarakat kurang mampu yang dimiliki oleh pemerintah kelurahan bersifat heterogen dan kompleks, mencakup berbagai aspek sosial-ekonomi seperti data pendapatan keluarga, pengeluaran bulanan, pekerjaan, jumlah tanggungan, dan tingkat pendidikan.

Untuk mengatasi permasalahan tersebut, diperlukan pendekatan berbasis teknologi informasi melalui penerapan data mining, khususnya teknik clustering (Komputer et al., 2024). Data mining memungkinkan ekstraksi informasi dan pengetahuan tersembunyi dari kumpulan data yang besar dan kompleks. Clustering merupakan teknik yang tepat untuk mengelompokkan masyarakat kurang mampu berdasarkan kemiripan karakteristik sosial-ekonomi, sehingga dapat dikategorikan

masyarakat mampu, kurang mampu, dan tidak mampu. Dengan pengelompokan yang sistematis, pihak kelurahan dan pemerintah daerah dapat menyusun strategi penyaluran bantuan yang lebih terarah dan mengalokasikan sumber daya bantuan sosial secara optimal.

Penelitian ini menggunakan algoritma OPTICS (*Ordering Points To Identify the Clustering Structure*) sebagai metode clustering untuk mengelompokkan masyarakat kurang mampu di Kelurahan Durian. Pemilihan algoritma OPTICS didasarkan pada beberapa pertimbangan. Pertama, data kemiskinan memiliki karakteristik heterogen dengan variasi kepadatan yang berbeda-beda, dimana terdapat kelompok masyarakat sangat miskin yang terkonsentrasi di area tertentu dan kelompok hampir miskin yang tersebar lebih jarang. OPTICS mampu mengidentifikasi cluster dengan variasi kepadatan berbeda dalam satu dataset. Kedua, OPTICS tidak memerlukan penentuan jumlah cluster di awal, sehingga lebih objektif dalam mengidentifikasi pola pengelompokan berdasarkan karakteristik data yang ada. Ketiga, OPTICS robust terhadap outlier atau data dengan kondisi ekstrem yang sering ditemukan dalam data sosial-ekonomi.

Proses klasterisasi menggunakan OPTICS dimulai dengan tahap preprocessing data yang meliputi data cleaning, normalisasi, dan transformasi data. Selanjutnya, algoritma OPTICS diterapkan dengan menentukan parameter epsilon (ϵ) dan MinPts, menghitung core distance dan reachability distance untuk setiap data point, membangun cluster-order, serta menghasilkan reachability plot yang menunjukkan struktur clustering dalam data. Dari cluster-order yang dihasilkan, dilakukan ekstraksi cluster untuk mengidentifikasi kelompok-kelompok

masyarakat kurang mampu berdasarkan tingkat kepadatan dan kemiripan karakteristik.

Hasil klasterisasi kemudian dievaluasi menggunakan metrik *Silhouette Index* untuk mengukur kualitas pengelompokan yang dihasilkan. *Silhouette Index* memberikan informasi tentang seberapa baik setiap objek telah ditempatkan dalam klasternya masing-masing serta seberapa jauh pemisahan antar klaster yang terbentuk. Untuk kelompok masyarakat kurang mampu di wilayah Kelurahan Durian.

Setelah cluster terbentuk dan dievaluasi, dilakukan analisis karakteristik setiap cluster untuk mengidentifikasi profil masing-masing kelompok (Dwitra Gusti Alriscki & Fauzan, 2024). Analisis ini mencakup statistik deskriptif dari variabel-variabel sosial-ekonomi pada setiap cluster, sehingga dapat diketahui pola dan ciri khas dari kelompok masyarakat mampu, kurang mampu, dan tidak mampu (Rady Putra & Anggrawan, 2021). Pemahaman mendalam tentang karakteristik setiap cluster menjadi dasar dalam merumuskan strategi penyaluran bantuan sosial yang tepat sasaran.

Kerangka berpikir penelitian ini menjelaskan bagaimana permasalahan ketidaktepatan penyaluran bantuan sosial di Kelurahan Durian Kota Medan dapat diatasi melalui penerapan algoritma clustering OPTICS (Rabbani et al., 2025). Dengan mengidentifikasi dan memetakan kelompok-kelompok masyarakat kurang mampu secara sistematis, penelitian ini diharapkan dapat memberikan rekomendasi kepada pihak kelurahan dan Dinas Sosial Kota Medan dalam merumuskan kebijakan penyaluran bantuan sosial yang lebih efektif dan efisien (Jhos Franklin Kemit, 2024). Output akhir dari penelitian ini adalah pengelompokan masyarakat

kurang mampu yang tervalidasi, visualisasi distribusi cluster, serta rekomendasi strategi penyaluran bantuan sosial yang disesuaikan dengan karakteristik dan tingkat kebutuhan masing-masing kelompok, sehingga dapat meningkatkan ketepatan sasaran dan dampak positif program bantuan sosial dalam upaya untuk mengurangi kemiskinan di Kelurahan Durian Kota Medan.

2.6 Ringkasan Penelitian Terdahulu

Untuk mendukung penelitian ini, dilakukan kajian terhadap beberapa penelitian terdahulu yang relevan dengan topik yang dibahas. Kajian ini bertujuan untuk melihat perkembangan penelitian sebelumnya sekaligus mengidentifikasi celah penelitian yang belum terpenuhi. Beberapa penelitian yang dijadikan acuan dirangkum dalam tabel berikut.

Tabel 2. 1 Penelitian Terdahulu

NO	Judul dan peneliti	Pembahasan	Metode	Kelebihan atau Kekurangan
1	Penerapan Metode OPTICS dan ST-DBSCAN Untuk Klasterisasi Data Kesehatan. Hastuti, S. H., Septiani, A. et al. (2024)	Penelitian ini mengelompokkan kecamatan di Kabupaten Lombok Timur berdasarkan jumlah dan distribusi tenaga kesehatan menggunakan metode OPTICS dan ST-DBSCAN, serta membandingkan keefektifan kedua metode tersebut dalam menangani data yang mengandung derau dan outlier untuk memberikan rekomendasi	Optics Dan St-DbScan	Kelebihan: memberikan kontribusi dalam penerapan algoritma OPTICS untuk klasterisasi data sosial-ekonomi kesehatan. Secara praktis, sehingga dapat menjadi dasar kebijakan distribusi tenaga kesehatan di Kabupaten Lombok Timur yang lebih tepat sasaran. Kekurangan: Kekurangan penelitian ini terletak pada penentuan parameter ST-

NO	Judul dan peneliti	Pembahasan	Metode	Kelebihan atau Kekurangan
		kebijakan penambahan tenaga kesehatan.		DBSCAN yang dilakukan secara manual tanpa penjelasan proses optimasi.
2	Klasterisasi Kabupaten/Kota Di Jawa Timur Berdasarkan Faktor Penyebab Perceraian Menggunakan Metode Optics (Ordering Points To Identify The Clustering Structure)	Penelitian ini mengelompokkan kabupaten/kota di Jawa Timur berdasarkan kemiripan distribusi faktor penyebab perceraian untuk mengidentifikasi pola dominan di setiap wilayah sebagai dasar perumusan kebijakan pencegahan perceraian yang tepat sasaran.	Optics	Kelebihan: memberikan kontribusi dalam penerapan algoritma OPTICS untuk klasterisasi data sosial-ekonomi kesehatan. Secara praktis, sehingga dapat menjadi dasar kebijakan distribusi tenaga kesehatan di Kabupaten Lombok Timur yang lebih tepat sasaran. Kekurangan: Dataset terlalu kecil (38 wilayah), data perceraian bersifat kategorikal/agregat yang tidak cocok untuk OPTICS yang dirancang untuk data numerik dengan variasi kepadatan, dan interpretasi core distance serta reachability distance sulit dijelaskan dalam konteks perceraian.
3	Implementasi Algoritma Clustering Dengan Modifikasi Metode Elbow Untuk Mendukung Strategi Pemerataan Bantuan Sosial Di Kabupaten Bojonegoro.	Penelitian ini Men Clustering 28 kecamatan di Bojonegoro berdasarkan 5 jenis bantuan sosial (2020-2021) untuk pemerataan distribusi bantuan. Modifikasi metode Elbow untuk menentukan k-cluster optimal.	K-Means & K-Medoids	Kelebihan: Perbandingan 2 metode (K-Means vs K-Medoids) Penentuan k-cluster optimal dengan Elbow Statistik deskriptif per cluster. Kekurangan: Data lebih besar (28 kecamatan, 5 variabel) Tidak ada

NO	Judul dan peneliti	Pembahasan	Metode	Kelebihan atau Kekurangan
	Hidayatul Fitriyah, Elsa Maulida Safitri, et al.			metrik evaluasi lain (Silhouette, DBI) Tidak menggunakan density-based method.
4	Penerapan Data Mining untuk Clustering Penduduk Miskin di Kota Tanjungbalai Menggunakan Metode Algoritma K-Means. Peneliti Zunaida Sitorus, Suhartika (2024)	Penelitian ini Men Clustering penduduk miskin di Tanjungbalai menggunakan K-Means untuk membantu Dinas Sosial dalam distribusi bantuan sosial. Data kemiskinan 2018-2022 menunjukkan tingkat 12,45%-14,64%.	K-Means	Kelebihan: Berbasis web, mudah diakses, Visualisasi clustering jelas Kekurangan: Sampel terlalu kecil (6 data), idak ada evaluasi cluster (Silhouette, DBI) Jumlah cluster ditentukan manual.

2.7 Analisis GAP

Berdasarkan tinjauan terhadap penelitian-penelitian relevan yang telah dilakukan sebelumnya, terdapat beberapa celah penelitian yang dapat diidentifikasi. Pertama, penelitian sebelumnya cenderung menggunakan algoritma partisi tunggal, khususnya K-Means dan K-Medoids, dalam mengelola data kemiskinan dan bantuan sosial. Penelitian di Kota Tanjungbalai dan Kabupaten Bojonegoro menggunakan K-Means yang memiliki keterbatasan signifikan yaitu sensitivitas terhadap inisialisasi centroid awal, ketergantungan pada penentuan jumlah cluster secara manual, serta tidak mampu mengidentifikasi outlier dan cluster dengan bentuk tidak beraturan. Sementara itu, penelitian yang menggunakan OPTICS masih terbatas pada konteks data kesehatan dan perceraian, belum diterapkan untuk kasus kemiskinan di Indonesia.

Kedua, belum ada penelitian yang secara spesifik menerapkan algoritma OPTICS dalam konteks klasterisasi masyarakat kurang mampu untuk optimalisasi penyaluran bantuan sosial di wilayah perkotaan Indonesia, khususnya Kota Medan. Penelitian yang membandingkan metode OPTICS dengan metode lain seperti ST-DBSCAN lebih fokus pada distribusi tenaga kesehatan, sehingga konteks dan karakteristik datanya berbeda dengan data kemiskinan yang bersifat multidimensional dan heterogen.

Ketiga, mayoritas penelitian dilakukan di wilayah dengan karakteristik geografis dan sosial-ekonomi yang berbeda dengan Kelurahan Durian Kota Medan. Kelurahan Durian sebagai salah satu wilayah dengan konsentrasi masyarakat kurang mampu yang cukup signifikan memiliki dinamika kemiskinan yang unik dan memerlukan analisis spesifik yang belum pernah dilakukan.

Keempat, evaluasi kualitas clustering pada penelitian terdahulu masih kurang komprehensif. Penelitian di Tanjungbalai tidak menggunakan metrik evaluasi cluster sama sekali, sedangkan penelitian di Bojonegoro hanya menggunakan Sum Squared Error (SSE) tanpa validasi menggunakan metrik standar seperti Silhouette Index.

Kelima, visualisasi hasil clustering pada penelitian sebelumnya masih terbatas pada scatter plot konvensional, tanpa memanfaatkan keunggulan spesifik dari algoritma yang digunakan seperti reachability plot pada OPTICS untuk mendukung pengambilan keputusan dalam penyaluran bantuan sosial.

Untuk mengisi celah penelitian yang ada, penelitian ini mengusulkan penerapan algoritma OPTICS (*Ordering Points To Identify the Clustering Structure*) pada wilayah Kelurahan Durian Kota Medan. Algoritma OPTICS dipilih

karena tidak memerlukan penentuan jumlah kluster di awal, sehingga pengelompokan masyarakat kurang mampu dapat dilakukan secara lebih objektif berdasarkan karakteristik data yang tersedia. Selain itu, OPTICS mampu mengenali kluster dengan tingkat kepadatan yang berbeda-beda dalam satu dataset, yang sangat sesuai dengan kondisi kemiskinan di lapangan yang bervariasi mulai dari kategori masyarakat mampu, kurang mampu, dan tidak mampu.

Keunggulan lain dari algoritma OPTICS adalah kemampuannya dalam mendeteksi dan menangani *outlier* secara otomatis, sehingga data dengan kondisi ekstrem tidak akan mempengaruhi kualitas hasil pengelompokan secara keseluruhan. Hasil dari proses ini juga divisualisasikan melalui *reachability plot*, yang dikombinasikan untuk memberikan gambaran yang lebih mudah dipahami dan komprehensif. Untuk memastikan kualitas pengelompokan yang dihasilkan, penelitian ini menggunakan *Silhouette Index* sebagai metrik evaluasi yang memberikan penilaian secara objektif.

Dengan demikian, penelitian ini tidak hanya bertujuan mengelompokkan masyarakat kurang mampu, tetapi juga memberikan kontribusi dari sisi metode dalam penerapan algoritma *clustering* berbasis densitas yang tepat guna, khususnya untuk mendukung optimalisasi penyaluran bantuan sosial di lingkungan perkotaan Indonesia, tepatnya di Kelurahan Durian Kota Medan.

BAB III

METODOLOGI PENELITIAN

3.1 Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif yang bersifat deskriptif-eksploratif dengan menggunakan pendekatan komputasional berbasis data. Penelitian ini bertujuan untuk memecahkan permasalahan nyata dalam proses identifikasi dan penyaluran bantuan sosial kepada masyarakat kurang mampu di Kelurahan Durian, Kota Medan, melalui penerapan teknik klusterisasi data. Metode yang diterapkan adalah metode eksperimen komputasi dengan memanfaatkan algoritma OPTICS (Ordering Points To Identify the Clustering Structure) untuk mengelompokkan data warga kurang mampu berdasarkan karakteristik kondisi sosial dan ekonomi mereka.

Berdasarkan tingkat penjelasannya, penelitian ini tergolong penelitian deskriptif karena menghasilkan gambaran yang sistematis, faktual, dan akurat mengenai karakteristik serta pola kelompok masyarakat kurang mampu melalui analisis data. Selain itu, penelitian ini juga merupakan penelitian terapan (applied research) karena bertujuan menghasilkan sistem pengelompokan berbasis data yang dapat digunakan secara langsung oleh pihak kelurahan dalam pengambilan keputusan terkait penyaluran bantuan sosial. Data yang digunakan bersumber dari data sekunder yang diperoleh dari Kelurahan Durian Kota Medan, berisi informasi sosial-ekonomi warga kurang mampu yang telah terdokumentasi dalam basis data kelurahan.

3.2 Teknik Pengumpulan Data

Penelitian ini menggunakan teknik pengumpulan data melalui studi lapangan yang dilaksanakan secara langsung di Kelurahan Durian, Kota Medan Timur, dengan terlebih dahulu melakukan koordinasi bersama aparaturnya guna memperoleh akses terhadap data sosial-ekonomi masyarakat kurang mampu yang telah terdokumentasi dalam sistem administrasi setempat. Data yang digunakan merupakan data sekunder yang bersumber dari basis data kelurahan serta daftar penerima program bantuan sosial, mencakup sejumlah variabel yang relevan dalam mengidentifikasi tingkat kemiskinan, antara lain pendapatan keluarga per bulan, pengeluaran bulanan, jenis pekerjaan kepala keluarga, jumlah tanggungan keluarga, dan tingkat pendidikan terakhir kepala keluarga, sebagaimana yang distandarkan oleh Badan Pusat Statistik untuk merepresentasikan kondisi sosial-ekonomi masyarakat secara komprehensif.

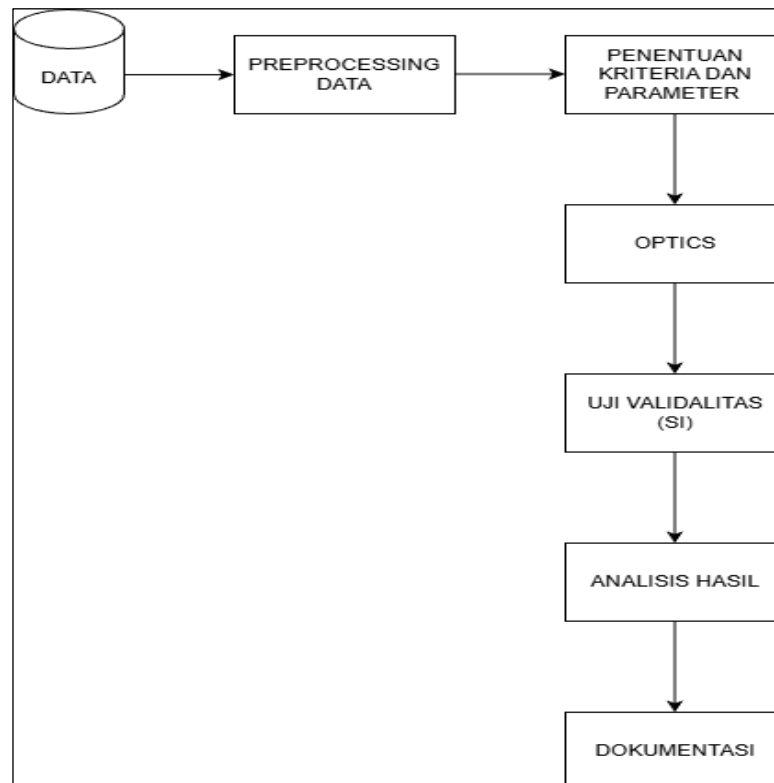
Di samping pengumpulan data lapangan, peneliti turut melaksanakan studi pustaka dengan mengkaji berbagai sumber akademik seperti jurnal ilmiah, buku referensi, laporan pemerintah, serta literatur lainnya yang berkaitan dengan klusterisasi data, algoritma OPTICS, dan permasalahan kemiskinan di Indonesia, guna memperoleh landasan teori, metodologi, serta rujukan penelitian terdahulu yang relevan.

No	Kode Penduduk	Pekerjaan	Pendapatan (Rp)	Pengeluaran (Rp)	Tingkat Pendidikan	Jml. Tanggungan
1	P01	Karyawan Swasta	5,000,000	1,500,000	D4/S1	2
2	P02	Karyawan Swasta	3,000,000	2,500,000	D1/D3	3
3	P03	Karyawan Swasta	3,500,000	1,000,000	SMA/Sederajat	0
4	P04	Karyawan Swasta	5,000,000	3,000,000	SMA/Sederajat	3
5	P05	Kontraktor Bangunan	2,000,000	1,000,000	S2/Sederajat	1
6	P06	Karyawan Swasta	1,500,000	1,000,000	SMA/Sederajat	2
7	P07	Wiraswasta	4,500,000	3,000,000	D4/S1	3
8	P08	Karyawan Swasta	2,500,000	1,500,000	SMA/Sederajat	1
9	P09	Buruh Bangunan	2,800,000	2,000,000	SMP/Sederajat	4
10	P10	Pedagang	2,000,000	1,800,000	SD/Sederajat	2
11	P11	Karyawan Swasta	5,500,000	4,000,000	D4/S1	2
12	P12	Pekerja Lepas	3,200,000	2,500,000	SMA/Sederajat	3
13	P13	PNS	6,500,000	4,500,000	S2/Sederajat	2
14	P14	Petani Sawit	2,200,000	1,900,000	SD/Sederajat	4
15	P15	Wiraswasta	8,000,000	5,500,000	D4/S1	3
16	P16	Buruh Bangunan	3,000,000	2,800,000	SMA/Sederajat	5
17	P17	Karyawan Swasta	4,800,000	3,500,000	D1/D3	2

Gambar 2. 2 Data Masyarakat Warga Kelurahan Durian

3.3 Arsitektur Penelitian

Arsitektur penelitian merupakan tahapan sistematis yang digunakan untuk menganalisis dan memproses data yang telah dikumpulkan dengan tujuan menemukan jawaban dari pertanyaan penelitian. Penelitian ini menggunakan pendekatan analisis data kuantitatif dengan algoritma OPTICS dalam mengklusterisasikan data masyarakat kurang mampu di Kelurahan Durian, Kota Medan Timur. Proses analisis dilakukan untuk mengelompokkan data berdasarkan karakteristik sosial-ekonomi dengan pendekatan berbasis kepadatan data. Adapun arsitektur penelitian dapat dilihat pada gambar berikut:



Gambar 3.1 Arsitektur Penelitian

3.3.1 Pengumpulan Data

Tahap ini mengumpulkan data masyarakat kurang mampu di Kelurahan Durian, Kota Medan Timur (Jhos Franklin Kemit, 2024). Data yang dikumpulkan mencakup informasi mengenai seperti data pendapatan, pengeluaran, pekerjaan,

tingkat pendidikan, dan jumlah tanggungan. pengumpulan data kemiskinan harus mencakup variabel-variabel yang dapat merepresentasikan kondisi sosial ekonomi masyarakat secara komprehensif.

No	Kode Penduduk	Pekerjaan	Pendapatan (Rp)	Pengeluaran (Rp)	Tingkat Pendidikan	Jml. Tanggungan
1	P01	Karyawan Swasta	5,000,000	1,500,000	D4/S1	2
2	P02	Karyawan Swasta	3,000,000	2,500,000	D1/D3	3
3	P03	Karyawan Swasta	3,500,000	1,000,000	SMA/Sederajat	0
4	P04	Karyawan Swasta	5,000,000	3,000,000	SMA/Sederajat	3
5	P05	Kontraktor Bangunan	2,000,000	1,000,000	S2/Sederajat	1
6	P06	Karyawan Swasta	1,500,000	1,000,000	SMA/Sederajat	2
7	P07	Wiraswasta	4,500,000	3,000,000	D4/S1	3
8	P08	Karyawan Swasta	2,500,000	1,500,000	SMA/Sederajat	1
9	P09	Buruh Bangunan	2,800,000	2,000,000	SMP/Sederajat	4
10	P10	Pedagang	2,000,000	1,800,000	SD/Sederajat	2
11	P11	Karyawan Swasta	5,500,000	4,000,000	D4/S1	2
12	P12	Pekerja Lepas	3,200,000	2,500,000	SMA/Sederajat	3
13	P13	PNS	6,500,000	4,500,000	S2/Sederajat	2
14	P14	Petani Sawit	2,200,000	1,900,000	SD/Sederajat	4
15	P15	Wiraswasta	8,000,000	5,500,000	D4/S1	3
16	P16	Buruh Bangunan	3,000,000	2,800,000	SMA/Sederajat	5
17	P17	Karyawan Swasta	4,800,000	3,500,000	D1/D3	2

Gambar 3.2 Data masyarakat kurang mampu

17	P17	Karyawan Swasta	4,800,000	3,500,000	D1/D3	2
18	P18	Pedagang	3,500,000	3,000,000	SMA/Sederajat	4
19	P19	PNS	7,200,000	5,000,000	D4/S1	3
20	P20	Pekerja Lepas	1,800,000	1,200,000	SMP/Sederajat	0
21	P21	Wiraswasta	5,400,000	4,200,000	SMA/Sederajat	2
22	P22	Karyawan Swasta	6,000,000	4,800,000	S2/Sederajat	1
23	P23	Petani Sawit	2,500,000	2,200,000	SD/Sederajat	3
24	P24	Buruh Bangunan	3,300,000	3,000,000	SMP/Sederajat	4
25	P25	Pedagang	2,900,000	2,600,000	SMA/Sederajat	2
26	P26	Wiraswasta	12,000,000	8,500,000	D4/S1	4
27	P27	Pekerja Lepas	2,700,000	2,500,000	SMA/Sederajat	2
28	P28	PNS	5,800,000	4,300,000	D4/S1	2
29	P29	Karyawan Swasta	4,500,000	3,400,000	SMA/Sederajat	3
30	P30	Buruh Bangunan	3,400,000	3,200,000	SMP/Sederajat	5
31	P31	Pedagang	4,000,000	3,300,000	SD/Sederajat	2
32	P32	Wiraswasta	10,500,000	7,600,000	D4/S1	4
33	P33	Petani Sawit	1,900,000	1,750,000	SMP/Sederajat	1
34	P34	Pekerja Lepas	2,300,000	1,300,000	SMA/Sederajat	0
35	P35	Karyawan Swasta	5,700,000	4,200,000	D1/D3	3

Gambar 3.3 Data masyarakat kurang mampu lanjut

3.3.2 Preprocessing Data

Setelah data dikumpulkan, tahap selanjutnya yaitu *preprocessing* untuk memastikan kualitas data sebelum data diolah. Tahapan *preprocessing* yang dilakukan dalam penelitian ini meliputi beberapa aktivitas sebagai berikut:

1. *Data cleaning* yaitu menghapus data yang tidak lengkap dan duplikatif agar dataset memiliki kualitas yang baik untuk dianalisis.

2. *Data integration* yaitu menggabungkan data dari tahun 2023-2025 menjadi satu dataset yang terintegrasi untuk analisis longitudinal.
3. *Data transformation* yaitu mengkonversikan data dengan format yang sesuai untuk proses analisis clustering.
4. *Data normalization* yaitu melakukan normalisasi dan standardisasi data untuk memastikan tidak ada variabel yang mendominasi dalam proses clustering karena perbedaan skala. Menurut Tang et al. (2021), normalisasi data sangat penting dalam algoritma density-based clustering agar atribut yang berbeda skala tidak mempengaruhi perhitungan jarak antar data point.

3.3.3 Penentuan Kriteria dan Parameter

Tahap ini melakukan penentuan kriteria dan parameter yang akan digunakan dalam proses clustering seperti berikut:

1. Kriteria clustering: Menentukan atribut dan fitur yang akan digunakan sebagai dasar pengelompokan masyarakat kurang mampu di Kelurahan Durian. Fitur-fitur yang digunakan dalam penelitian ini adalah:
 - a. Pendapatan Keluarga: Total pendapatan keluarga per bulan
 - b. Pengeluaran Bulanan: Total pengeluaran keluarga per bulan
 - c. Pekerjaan Keluarga: Jenis pekerjaan yang dimiliki oleh kepala keluarga atau anggota keluarga.
 - d. Jumlah Tanggungan Keluarga: Banyaknya anggota keluarga yang menjadi tanggungan.
 - e. Tingkat Pendidikan Keluarga: Jenjang pendidikan terakhir yang dikonversi menjadi nilai numerik

Fitur-fitur ini dipilih berdasarkan standar indikator kemiskinan BPS dan akan dinormalisasi menggunakan Min-Max Scaling sebelum diproses.

2. Parameter algoritma: Parameter algoritma OPTICS yang perlu ditentukan adalah nilai epsilon (ϵ) dan MinPts.

a. Parameter min_samples (MinPts): Parameter min_samples adalah jumlah minimum point dalam ϵ -neighborhood agar suatu point dianggap sebagai core point. Nilai yang digunakan berdasarkan dimensi data (5 fitur), yaitu $2 \times 5 = 10$. Penelitian ini akan melakukan eksperimen dengan nilai 5, 10, 15, dan 20 untuk menemukan nilai optimal.

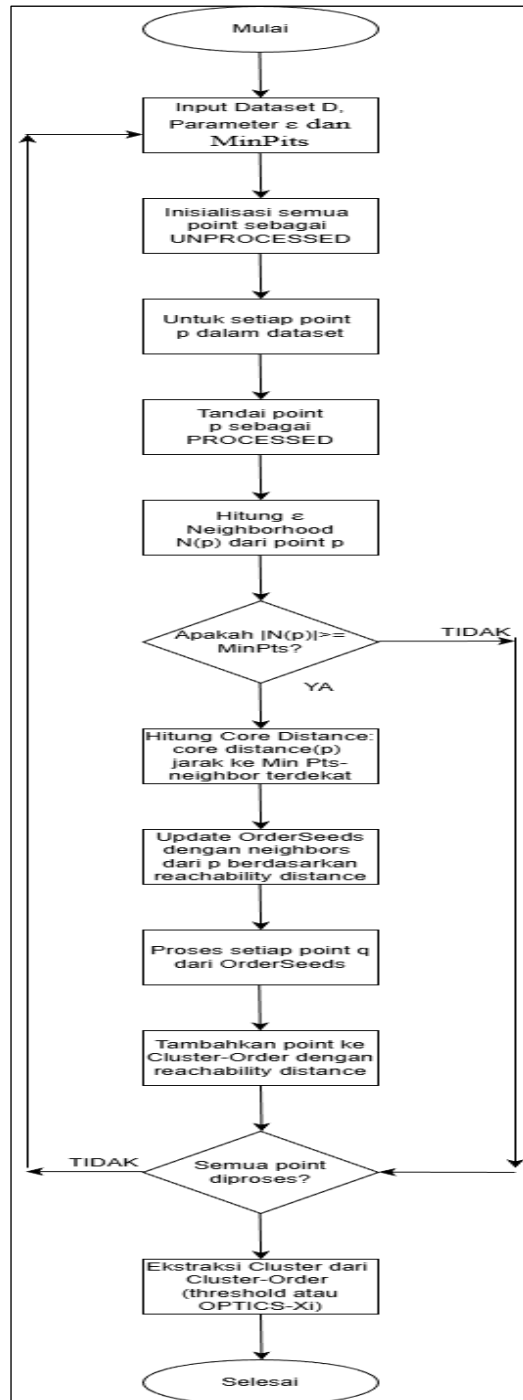
b. Parameter max_eps (ϵ / Epsilon)

Parameter max_eps adalah jarak maksimum untuk mendefinisikan neighborhood dari suatu point. Penentuan nilai dilakukan melalui analisis k-distance plot dan eksperimen dengan nilai 0.3, 0.5, 0.7, dan 1.0 untuk data yang telah dinormalisasi. Nilai optimal dipilih berdasarkan Silhouette Index tertinggi.

Selain itu, proses evaluasi clustering juga mempertimbangkan kestabilan jumlah cluster yang terbentuk pada setiap kombinasi parameter yang diuji.

3.3.4 Proses dengan Algoritma *OPTICS*

Algoritma OPTICS (Ordering Points To Identify the Clustering Structure) digunakan untuk mengelompokkan data masyarakat kurang mampu berdasarkan kepadatan karakteristik data.



Gambar 3. 4 Flowchart OPTICS

Proses algoritma ini berjalan sebagai berikut:

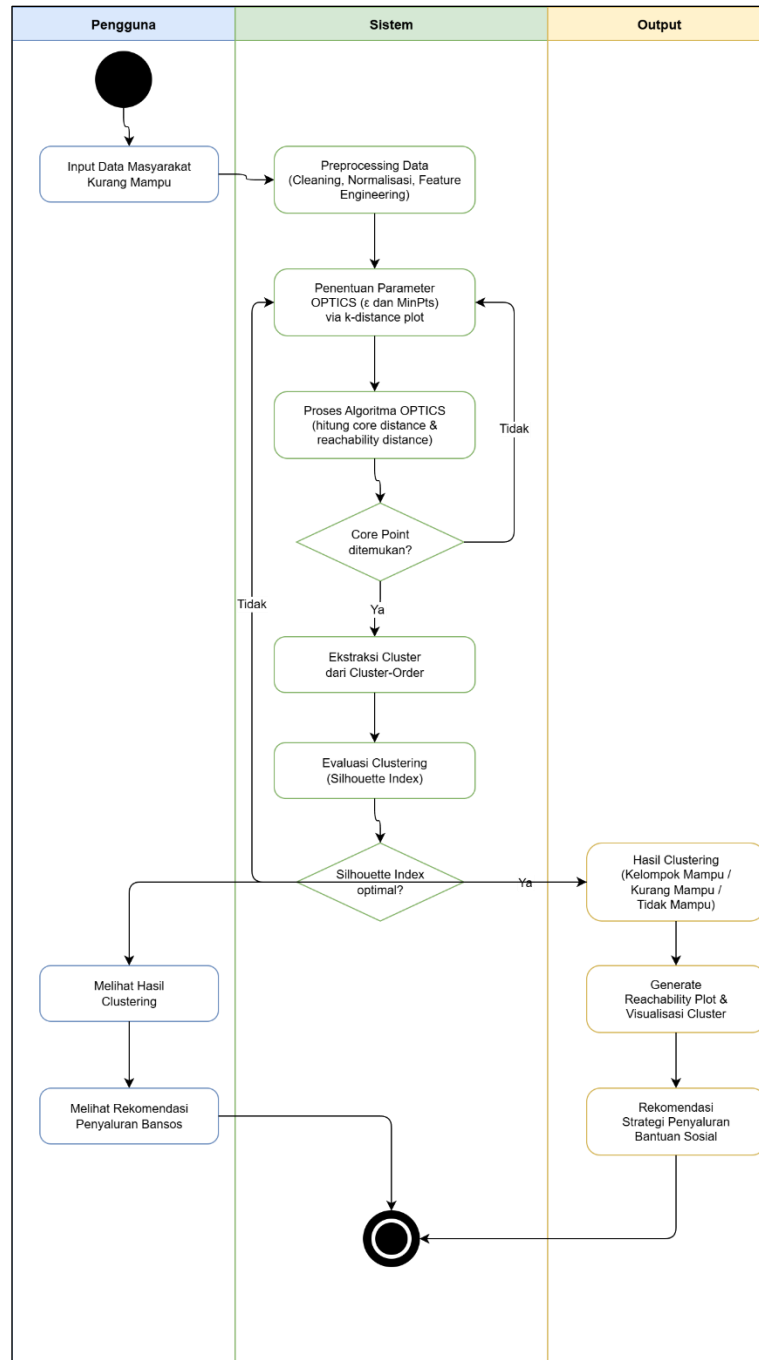
1. Inisialisasi: Proses dimulai dengan input dataset D yang berisi data masyarakat kurang mampu serta parameter ϵ (epsilon) dan MinPts. Seluruh point dalam

dataset diberi status 'UNPROCESSED' sebagai penanda bahwa point tersebut belum diproses.

2. Iterasi Setiap Point: Algoritma melakukan iterasi untuk setiap point p dalam dataset. Point p kemudian ditandai sebagai 'PROCESSED' dan dihitung ϵ -Neighborhood $N(p)$, yaitu semua point yang berada dalam radius ϵ dari point p .
3. Pengecekan Core Point: Dilakukan pengecekan apakah jumlah point dalam $N(p)$ lebih besar atau sama dengan MinPts ($|N(p)| \geq \text{MinPts}$). Jika ya, maka point p merupakan core point dan algoritma melanjutkan ke perhitungan core distance. Jika tidak, point p dianggap sebagai noise atau border point.
4. Perhitungan Core Distance: Untuk point yang merupakan core point, dihitung core distance yaitu jarak dari point p ke MinPts -neighbor terdekat. Nilai ini merepresentasikan tingkat kepadatan lokal di sekitar point p .
5. Update OrderSeeds: OrderSeeds adalah priority queue yang menyimpan neighbors dari point p berdasarkan reachability distance. Reachability distance dihitung sebagai nilai maksimum antara core distance point p dan jarak Euclidean antara p dengan neighbor-nya.
6. Pemrosesan OrderSeeds: Setiap point q dari OrderSeeds diproses secara berurutan berdasarkan reachability distance terkecil. Point q kemudian ditambahkan ke Cluster-Order bersama dengan nilai reachability distance-nya, yang akan digunakan untuk visualisasi dan ekstraksi cluster.
7. Iterasi Hingga Selesai: Proses iterasi berlanjut hingga semua point dalam dataset telah diproses. Hasil akhir berupa Cluster-Order yang merepresentasikan struktur clustering berbasis kepadatan.

8. Ekstraksi Cluster: Setelah Cluster-Order terbentuk, dilakukan ekstraksi cluster menggunakan metode threshold atau algoritma OPTICS untuk mengidentifikasi kelompok-kelompok masyarakat kurang mampu berdasarkan pola kepadatan yang terbentuk.

3.3.5 Activity Diagram



Gambar 3.5 Activity Diagram

Activity diagram ini menggambarkan alur proses sistem dalam melakukan pengelompokan (clustering) data masyarakat kurang mampu menggunakan algoritma OPTICS. Proses dimulai dari pengguna yang memasukkan data masyarakat kurang mampu ke dalam sistem. Data yang telah diinput kemudian diproses oleh sistem melalui tahap preprocessing, yang meliputi *data cleaning* untuk menghilangkan data yang tidak valid atau duplikat, *normalisasi* untuk menyamakan skala data, serta *feature engineering* untuk membentuk fitur yang lebih representatif.

Setelah preprocessing selesai, sistem melakukan penentuan parameter OPTICS, yaitu nilai *epsilon* (ϵ) dan *minimum points* (*MinPts*). Penentuan parameter ini dilakukan menggunakan metode *k-distance plot* agar menghasilkan nilai yang optimal untuk proses clustering.

Selanjutnya, sistem menjalankan algoritma OPTICS dengan menghitung *core distance* dan *reachability distance* untuk setiap data. Berdasarkan hasil perhitungan tersebut, sistem akan mengecek apakah terdapat core point. Jika tidak ditemukan, maka proses akan kembali untuk penyesuaian parameter. Jika ditemukan, maka proses dilanjutkan ke tahap berikutnya.

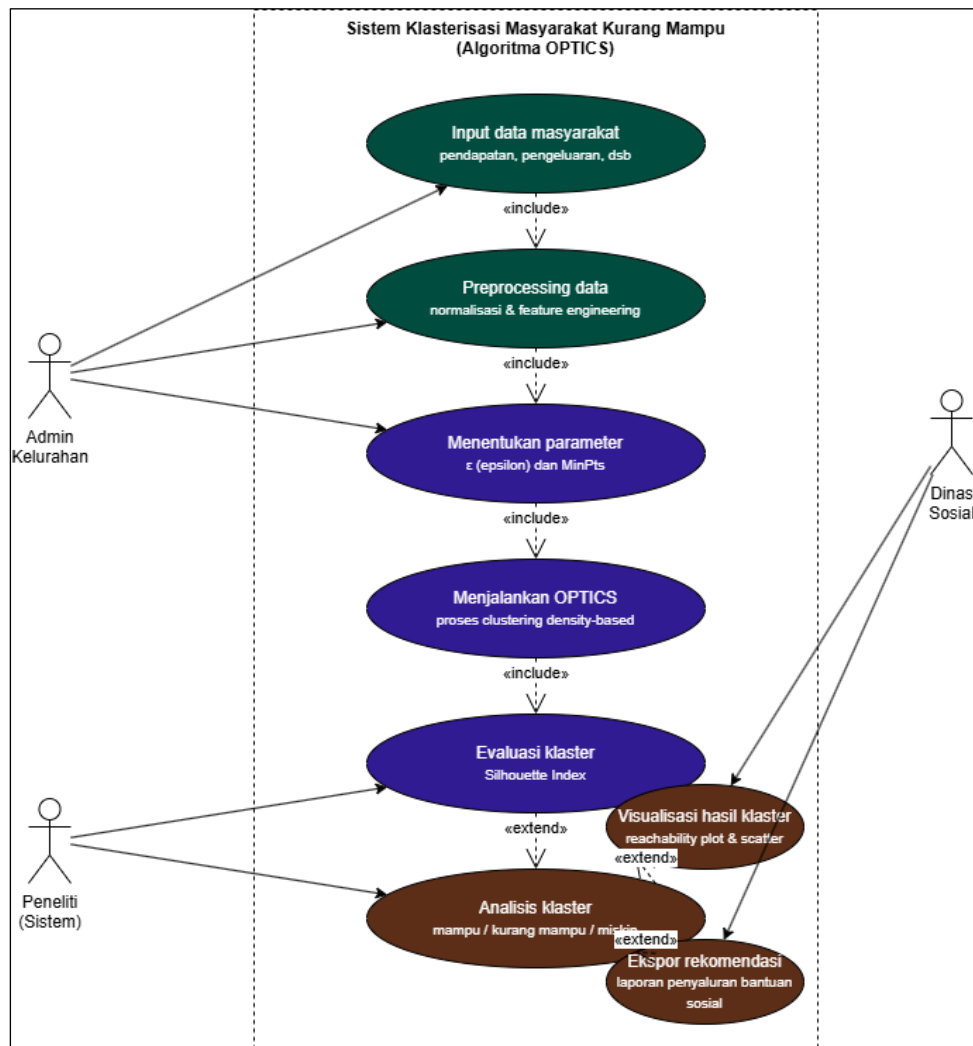
Pada tahap selanjutnya, sistem melakukan ekstraksi cluster berdasarkan urutan cluster (*cluster-order*) yang dihasilkan oleh algoritma OPTICS. Setelah cluster terbentuk, dilakukan evaluasi hasil clustering menggunakan *Silhouette Index* untuk mengukur kualitas pengelompokan.

Jika nilai Silhouette Index belum optimal, maka sistem akan kembali ke tahap sebelumnya untuk penyesuaian parameter. Namun, jika sudah optimal, maka

sistem akan menghasilkan output clustering, yaitu pengelompokan masyarakat ke dalam kategori seperti *mampu*, *kurang mampu*, dan *tidak mampu*.

Selain itu, sistem juga menghasilkan visualisasi dalam bentuk reachability plot dan grafik cluster untuk mempermudah interpretasi hasil. Pengguna kemudian dapat melihat hasil clustering tersebut serta memperoleh rekomendasi strategi penyaluran bantuan sosial berdasarkan hasil pengelompokan yang telah dilakukan. Proses berakhir setelah pengguna mendapatkan hasil analisis dan rekomendasi dari sistem.

3.3.6 Use Case Diagram



Gambar 3.6 Use Case Diagram

Diagram ini menggambarkan alur proses sistem dalam melakukan analisis dan pengelompokan data masyarakat berdasarkan tingkat kesejahteraan menggunakan algoritma OPTICS.

Proses dimulai dari input data masyarakat, yang mencakup berbagai atribut seperti pendapatan, pengeluaran, dan variabel pendukung lainnya. Data yang telah dimasukkan kemudian diproses pada tahap preprocessing data, yang meliputi normalisasi dan *feature engineering* untuk memastikan data berada dalam skala yang seragam dan memiliki representasi fitur yang optimal.

Selanjutnya, sistem melakukan penentuan parameter OPTICS, yaitu nilai *epsilon* (ϵ) dan *MinPts*. Parameter ini sangat penting karena akan mempengaruhi hasil pembentukan cluster pada proses berikutnya.

Setelah parameter ditentukan, sistem menjalankan algoritma OPTICS, yaitu metode clustering berbasis kepadatan (*density-based clustering*) yang mampu mengidentifikasi struktur cluster berdasarkan kerapatan data.

Hasil clustering yang diperoleh kemudian dievaluasi pada tahap evaluasi kluster menggunakan metode *Silhouette Index* untuk mengukur kualitas pengelompokan data. Pada tahap ini juga terdapat proses tambahan (*extend*), yaitu visualisasi hasil kluster dalam bentuk *reachability plot* dan *scatter plot*, yang membantu dalam memahami struktur cluster secara visual.

Selanjutnya, sistem melakukan analisis kluster, di mana hasil pengelompokan diinterpretasikan menjadi kategori seperti masyarakat mampu, kurang mampu, dan miskin. Dari hasil analisis ini, sistem juga menghasilkan proses lanjutan (*extend*) berupa ekspor rekomendasi, yaitu laporan strategi penyaluran bantuan sosial yang dapat digunakan sebagai dasar pengambilan keputusan.

Secara keseluruhan, diagram ini menunjukkan hubungan antar proses yang saling terintegrasi, di mana setiap tahapan utama saling terhubung melalui relasi *include*, serta didukung oleh proses tambahan (*extend*) berupa visualisasi dan rekomendasi untuk memperkaya hasil analisis.

3.3.7 Analisis hasil dan perbandingan

Setelah proses clustering menggunakan algoritma OPTICS selesai dilakukan, tahap selanjutnya adalah melakukan analisis terhadap hasil clustering dan perbandingan antar cluster yang terbentuk. Analisis dilakukan dengan mengevaluasi kualitas clustering menggunakan metrik Silhouette Index yang mengukur seberapa baik setiap data point ditempatkan dalam cluster-nya dan seberapa terpisah cluster-cluster yang terbentuk, dengan nilai berkisar dari -1 hingga 1 di mana nilai mendekati 1 menunjukkan clustering yang baik.

Selanjutnya dilakukan analisis karakteristik setiap cluster dengan menghitung statistik deskriptif seperti nilai rata-rata, median, minimum, dan maksimum dari setiap variabel sosial-ekonomi (pendapatan, pengeluaran, pekerjaan, tingkat pendidikan, dan jumlah tanggungan) pada masing-masing cluster yang terbentuk. Perbandingan antar cluster dilakukan untuk mengidentifikasi perbedaan signifikan antara kelompok masyarakat mampu, kurang mampu, dan tidak mampu berdasarkan profil karakteristik sosial-ekonomi yang dimiliki setiap cluster.

Hasil analisis ini kemudian divisualisasikan melalui reachability plot untuk menunjukkan struktur clustering berbasis kepadatan, untuk memvisualisasikan distribusi cluster untuk menunjukkan sebaran kelompok masyarakat kurang mampu di wilayah Kelurahan Durian. Perbandingan hasil clustering juga dilakukan dengan

melakukan eksperimen menggunakan beberapa kombinasi parameter ϵ dan MinPts untuk menemukan konfigurasi optimal yang menghasilkan Silhouette Index tertinggi dan cluster yang paling bermakna secara interpretatif. Analisis dan perbandingan ini menjadi dasar dalam merumuskan rekomendasi strategi penyaluran bantuan sosial yang tepat sasaran sesuai dengan karakteristik dan tingkat kebutuhan masing-masing kelompok masyarakat kurang mampu di Kelurahan Durian Kota Medan.

BAB IV

HASIL DAN PEMBAHASAN

4.1. Deskripsi data

Pengumpulan data dilakukan melalui studi lapangan langsung di Kelurahan Durian, Kota Medan Timur. Data yang digunakan merupakan data sekunder yang telah terdokumentasi dalam catatan administrasi kelurahan mengenai masyarakat kurang mampu. Data diperoleh dalam bentuk daftar warga yang mencakup informasi sosial-ekonomi dari setiap kepala keluarga atau anggota keluarga yang terdaftar sebagai masyarakat kurang mampu di wilayah tersebut.

Dataset yang digunakan dalam penelitian ini terdiri dari 800 data Masyarakat kelurahan Durian dengan 5 variabel utama yang mencerminkan kondisi sosial-ekonomi masyarakat, yaitu: (1) pendapatan keluarga per bulan, (2) pengeluaran bulanan, (3) jenis pekerjaan, (4) tingkat pendidikan terakhir, dan (5) jumlah tanggungan keluarga. Kelima variabel ini dipilih berdasarkan standar indikator kemiskinan Badan Pusat Statistik (BPS) dan relevansinya dalam proses klusterisasi berbasis kepadatan.

Tabel 4.1 Statistik Data Masyarakat Kelurahan Durian

Variabel	Minimum	Maksimum	Rata-rata	Std. Deviasi
Pendapatan (Rp)	Rp 0	Rp 21.000.000	Rp 4.250.000	Rp 3.120.000
Pengeluaran (Rp)	Rp 0	Rp 20.000.000	Rp 3.570.000	Rp 2.140.000
Jumlah Tanggungan	0	6	2,29	1,42

Berdasarkan Tabel 4.1, rata-rata pendapatan masyarakat adalah Rp 4.250.000 per bulan dengan nilai minimum Rp 0 (tidak memiliki penghasilan) dan nilai maksimum Rp 21.000.000. Terdapat sejumlah masyarakat yang tercatat dengan pendapatan Rp 0, mengindikasikan bahwa mereka sepenuhnya tidak

bekerja atau bergantung pada anggota keluarga lain. Rata-rata pengeluaran sebesar Rp 3.570.000 dengan standar deviasi yang cukup tinggi, mencerminkan heterogenitas kondisi ekonomi masyarakat yang menjadi objek penelitian.

Tabel 4.2 Distribusi Jenis Pekerjaan Masyarakat Kelurahan Durian

No	Jenis Pekerjaan	Jumlah	Persentase (%)
1	Karyawan Swasta	~267	33,4
2	Tidak Bekerja	~189	23,6
3	Wiraswasta	~78	9,7
4	Freelance	~55	6,9
5	Driver Taxi Online	~34	4,2
6	Petani	~22	2,8
7	Lainnya	~155	19,4
Total		800	100

Berdasarkan Tabel 4.2, jenis pekerjaan terbanyak adalah Karyawan Swasta (sekitar 33,4%), diikuti oleh Tidak Bekerja (sekitar 23,6%). Tingginya proporsi masyarakat yang tidak bekerja menunjukkan tingkat ketergantungan ekonomi yang signifikan di Kelurahan Durian dan menjadi indikator penting dalam proses klusterisasi kemiskinan.

Tabel 4.3 Distribusi Tingkat Pendidikan Masyarakat Kelurahan Durian

No	Tingkat Pendidikan	Jumlah	Persentase (%)
1	SMA/Sederajat	~445	55,6
2	D4/S1	~145	18,1
3	SD/Sederajat	~89	11,1
4	D1/D3	~78	9,7
5	SMP/Sederajat	~43	5,5
Total		800	100

4.2. Preprocessing Data

Sebelum data diproses menggunakan algoritma OPTICS, dilakukan tahap preprocessing untuk memastikan kualitas dan konsistensi data. Tahap preprocessing meliputi beberapa langkah sistematis sebagai berikut.

4.2.1. Data Cleaning

Pemeriksaan data dilakukan untuk mengidentifikasi nilai kosong (missing values), data duplikat, dan inkonsistensi format. Dari 800 data Masyarakat kelurahan durian yang dikumpulkan, seluruh data teridentifikasi dalam kondisi lengkap untuk kelima variabel yang digunakan. Tidak ditemukan nilai kosong maupun data duplikat, sehingga tidak diperlukan penghapusan atau imputasi data.

4.2.2. Penghapusan Outlier

Penghapusan outlier dilakukan menggunakan metode Z-Score dengan threshold 3 standar deviasi. Dari 800 data awal, terdapat 11 baris yang teridentifikasi sebagai outlier ekstrem dan dihapus dari dataset, sehingga tersisa 789 data yang akan digunakan dalam proses clustering. Rata-rata pendapatan setelah preprocessing menjadi Rp 4,12 juta dengan range Rp 0 – Rp 12,60 juta, lebih terkontrol dibandingkan sebelumnya yang mencapai Rp 21,00 juta.

4.2.3. Encoding Variabel Kategorikal

Variabel Pekerjaan dan Tingkat Pendidikan merupakan variabel bertipe kategorikal yang harus dikonversi ke nilai numerik sebelum dapat diproses oleh algoritma OPTICS. Konversi dilakukan menggunakan ordinal encoding berdasarkan tingkat sosial-ekonomi masing-masing kategori.

Tabel 4.4 Encoding Tingkat Pendidikan

Tingkat Pendidikan	Nilai Numerik
Tidak Sekolah	0
SD/Sederajat	1
SMP/Sederajat	2
SMA/Sederajat	3
D1/D3	4
D4/S1	5
S2/S3	6

Tabel 4.4 menunjukkan proses encoding tingkat pendidikan responden ke dalam bentuk numerik. Setiap kategori pendidikan diberi nilai berdasarkan jenjangnya, dimulai dari tidak sekolah dengan nilai 0 hingga pendidikan S2/S3 dengan nilai 6. Pemberian nilai ini bertujuan untuk memudahkan proses analisis data secara kuantitatif.

Tabel 4.5 Encoding Jenis Pekerjaan

Jenis Pekerjaan	Nilai Numerik
Tidak Bekerja	0
Buruh / Buruh Bangunan	1
Petani / Nelayan	2
Pedagang / Supir / Driver / Musisi	3
Freelance / Sales / Kontraktor / Ahli Gigi	4
Wiraswasta / Karyawan Swasta / Pemilik Kos	5
Pegawai Negeri Sipil	6
Dosen	7

Tabel 4.5 menyajikan encoding jenis pekerjaan responden ke dalam nilai numerik. Kategori pekerjaan dikelompokkan dan diberikan nilai secara bertahap, dimulai dari tidak bekerja dengan nilai 0 hingga pekerjaan dengan tingkat keahlian atau stabilitas yang lebih tinggi dengan nilai yang lebih besar. Transformasi ini dilakukan agar data kategorikal dapat diolah dalam metode analisis berbasis numerik

4.2.4. Feature Engineering

Feature engineering dilakukan untuk memperkaya representasi data sebelum clustering. Fitur-fitur tambahan yang dibuat meliputi: (1) rasio pendapatan terhadap pengeluaran (Income_Expense_Ratio), (2) pendapatan per tanggungan (Income_Per_Dependent), (3) transformasi logaritmik untuk variabel berdistribusi skewed (Log_Pendapatan dan Log_Pengeluaran), (4) binning jumlah tanggungan, (5) interaksi antara fitur pendidikan dan pekerjaan, (6) surplus/defisit keuangan, serta (7) mean encoding dan count encoding untuk variabel kategorikal.

4.2.5. Normalisasi Data

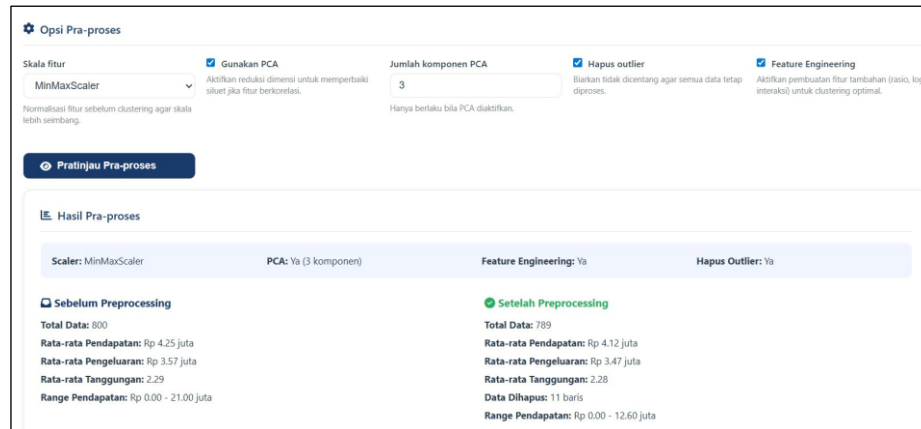
Normalisasi dilakukan menggunakan metode Min-Max Scaling untuk menyamakan skala seluruh variabel ke rentang [0, 1]. Hal ini penting dalam algoritma berbasis jarak seperti OPTICS, agar variabel dengan skala besar (seperti pendapatan dalam satuan rupiah) tidak mendominasi perhitungan jarak Euclidean dibandingkan variabel dengan skala kecil (seperti jumlah tanggungan).

Setelah normalisasi, seluruh nilai fitur berada dalam rentang 0 hingga 1, sehingga perhitungan jarak antar data point menjadi proporsional dan tidak bias terhadap skala pengukuran masing-masing variabel.

Tabel 4.6 Ringkasan Hasil Preprocessing Data

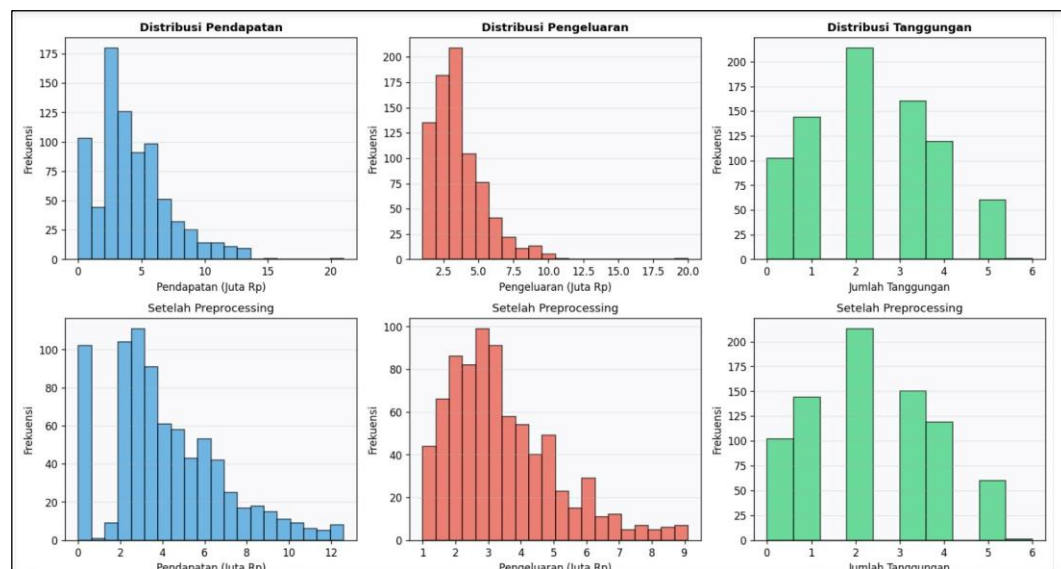
Keterangan	Sebelum Preprocessing	Setelah Preprocessing
Total Data	800 baris	789 baris
Rata-rata Pendapatan	Rp 4,25 juta	Rp 4,12 juta
Rata-rata Pengeluaran	Rp 3,57 juta	Rp 3,47 juta
Rata-rata Tanggungan	2,29	2,28
Data Dihapus (Outlier)	-	11 baris
Range Pendapatan	Rp 0 – Rp 21,00 juta	Rp 0 – Rp 12,60 juta
Scaler	-	MinMaxScaler
PCA	-	Ya (3 komponen)
Feature Engineering	-	Ya

Gambar 4.1 menunjukkan hasil pratinjau pra-proses pada aplikasi, termasuk perbandingan statistik sebelum dan sesudah preprocessing. Gambar 4.2 menampilkan distribusi masing-masing variabel sebelum dan sesudah preprocessing, sedangkan Gambar 4.3 menampilkan 10 data pertama setelah preprocessing.



Gambar 4.1 Hasil Pratinjau Pra-Proses pada Aplikasi

Berdasarkan Gambar 4.1 menampilkan hasil pratinjau proses pra-pemrosesan data pada aplikasi. Pada tahap ini dilakukan beberapa langkah seperti normalisasi data menggunakan MinMaxScaler, penerapan PCA (Principal Component Analysis), serta penghapusan outlier. Hasilnya menunjukkan adanya perubahan pada jumlah data dan nilai statistik, sehingga data menjadi lebih siap untuk tahap analisis selanjutnya.



Gambar 4.2 Distribusi Variabel Sebelum dan Sesudah Preprocessing

Berdasarkan Gambar 4.2 menunjukkan distribusi variabel sebelum dan sesudah preprocessing, meliputi pendapatan, pengeluaran, dan jumlah tanggungan.

Terlihat bahwa setelah preprocessing, distribusi data menjadi lebih merata dan outlier berkurang, sehingga pola data lebih mudah dianalisis dan tidak terlalu bias.

Preview 10 Data Pertama (Setelah Preprocessing)						
No	Kode	Pekerjaan	Pendapatan (Rp)	Pengeluaran (Rp)	Pendidikan	Tanggung
1	R01	Karyawan Swasta	Rp 5.000.000	Rp 1.500.000	D4/S1	2
2	R02	Karyawan Swasta	Rp 1.500.000	Rp 2.000.000	SMA/Sederajat	0
3	R03	Tidak Bekerja	Rp 0	Rp 1.000.000	SMP/Sederajat	0
4	R04	Karyawan Swasta	Rp 2.000.000	Rp 3.000.000	SMA/Sederajat	3
5	R05	Karyawan Swasta	Rp 3.000.000	Rp 2.500.000	D1/D3	3
6	R06	Tidak Bekerja	Rp 0	Rp 1.000.000	SMA/Sederajat	1
7	R07	Karyawan Swasta	Rp 3.500.000	Rp 1.000.000	SMA/Sederajat	0
8	R08	Karyawan Swasta	Rp 5.000.000	Rp 3.000.000	SMA/Sederajat	3
9	R09	Kontraktor	Rp 2.000.000	Rp 1.000.000	S2/Sederajat	1
10	R10	Karyawan Swasta	Rp 1.500.000	Rp 1.000.000	SMA/Sederajat	2

Gambar 4.3 Preview 10 Data Pertama Setelah Preprocessing

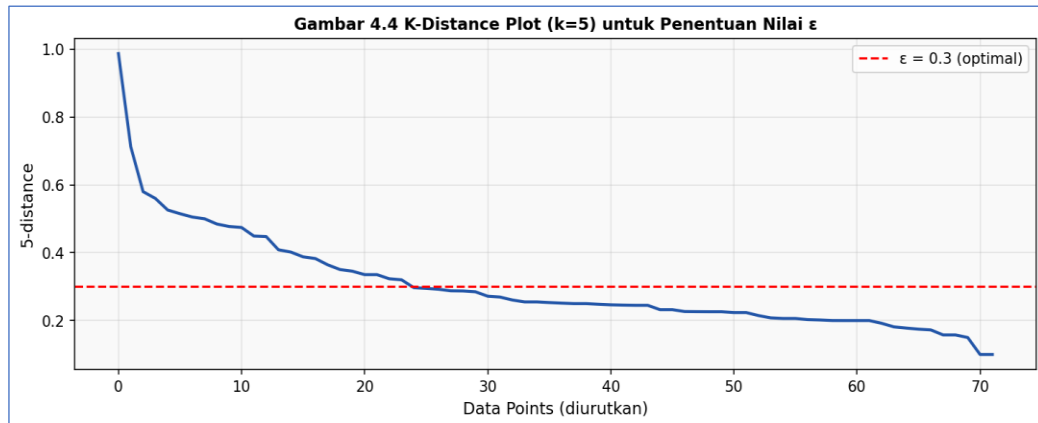
Berdasarkan Gambar 4.3 menyajikan preview 10 data pertama setelah dilakukan preprocessing. Data yang ditampilkan telah melalui tahap pembersihan dan transformasi, sehingga lebih terstruktur dan siap digunakan dalam proses pemodelan atau analisis lebih lanjut.

4.3. Penentuan Parameter Optics

Penentuan parameter optimal untuk algoritma OPTICS dilakukan melalui dua tahapan, yaitu analisis K-Distance Plot untuk menentukan nilai epsilon (ϵ) dan eksperimen grid search untuk menemukan kombinasi parameter `min_samples` (MinPts), `max_eps`, dan `xi` (ξ) yang menghasilkan Silhouette Index tertinggi.

4.3.1. Analisis K-Distance Plot

K-Distance Plot digunakan sebagai panduan visual untuk menentukan nilai epsilon (ϵ) yang sesuai. Grafik ini menampilkan jarak ke tetangga ke-k dari setiap data point, diurutkan secara menurun. Nilai ϵ yang optimal berada pada titik "siku" (elbow) dari grafik, yaitu titik di mana terjadi perubahan kemiringan yang signifikan.

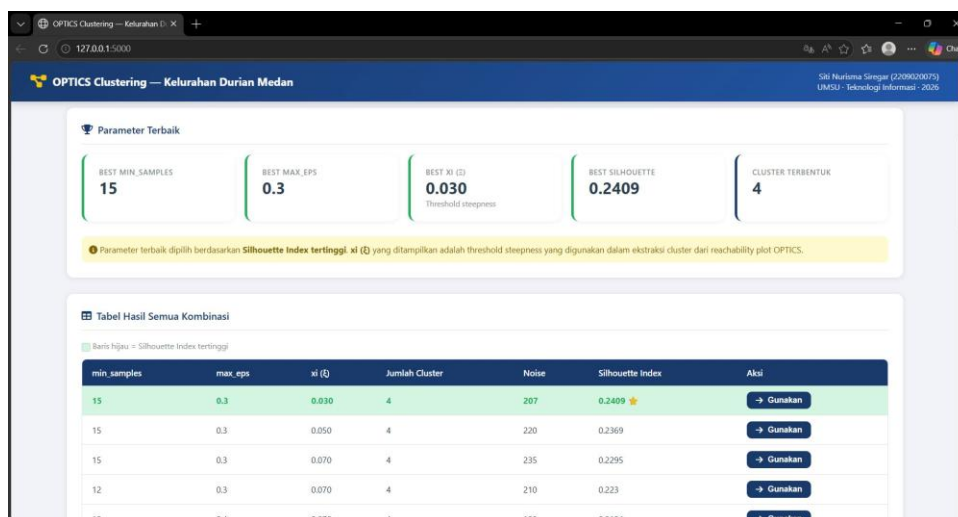


Gambar 4.4 K - Distance Plot untuk Penentuan Nilai Optimal ϵ

Berdasarkan Gambar 4.4, titik siku teridentifikasi pada nilai jarak sekitar 0,3. Oleh karena itu, nilai $\max_eps = 0,3$ ditetapkan sebagai kandidat utama untuk diuji lebih lanjut melalui proses grid search.

4.3.2. Grid Search Parameter

Eksperimen grid search dilakukan dengan menguji berbagai kombinasi nilai $\min_samples$ dan \max_eps . Kualitas setiap kombinasi dievaluasi menggunakan Silhouette Index. Aplikasi secara otomatis menampilkan parameter terbaik berdasarkan Silhouette Index tertinggi, sebagaimana ditunjukkan pada Gambar 4.5.



Gambar 4.5 Hasil Auto Tune Parameter dan Tabel Grid Search OPTICS

Berdasarkan Tabel 4.7, kombinasi parameter $\text{min_samples} = 15$, $\text{max_eps} = 0,3$, dan $\text{xi} (\xi) = 0,030$ menghasilkan Silhouette Index tertinggi sebesar 0,2409 dengan membentuk 4 cluster dan 207 data noise. Parameter ini ditetapkan sebagai konfigurasi optimal untuk proses klasterisasi selanjutnya. Nilai xi yang ditampilkan merupakan threshold steepness yang digunakan dalam ekstraksi cluster dari reachability plot OPTICS.

Tabel 4.7 Hasil Grid Search Parameter OPTICS (Kombinasi Terbaik)

min_samples	max_eps	xi (ξ)	Jumlah Cluster	Noise	Silhouette Index
15	0,3	0,030	4	207	0,2409 ★
15	0,3	0,050	4	220	0,2369
15	0,3	0,070	4	235	0,2295
12	0,3	0,070	4	210	0,223

4.4. Proses Algoritma OPTICS

Setelah parameter optimal ditetapkan, algoritma OPTICS dijalankan pada dataset yang telah melalui preprocessing (789 data). Proses OPTICS bekerja dengan membangun cluster-ordering berdasarkan reachability distance setiap data point, menggunakan parameter $\text{min_samples} = 15$, $\text{max_eps} = 0,3$, dan $\text{xi} = 0,030$.

4.4.1. Perhitungan Core Distance dan Reachability Distance

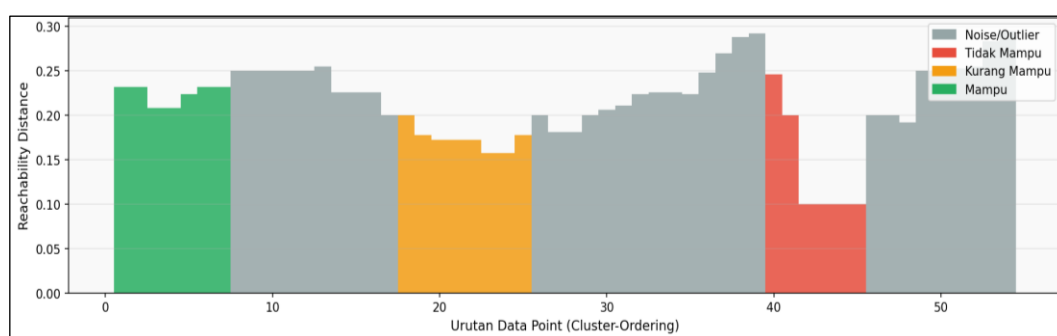
Untuk setiap data point p dalam dataset, algoritma menghitung core distance sebagai jarak ke tetangga ke- MinPts terdekat dalam radius ϵ . Suatu titik dianggap sebagai core point apabila memiliki minimal 15 tetangga dalam radius 0,3 (setelah normalisasi).

Titik-titik yang tidak memenuhi syarat sebagai core point dan tidak dapat dijangkau dari core point manapun dalam radius ϵ dikategorikan sebagai noise

(outlier). Dalam hasil penelitian ini, terdapat 207 data point yang dikategorikan sebagai noise.

4.4.2. Reachability Plot

Reachability plot merupakan visualisasi utama yang dihasilkan oleh algoritma OPTICS. Grafik ini menampilkan urutan data point pada sumbu x dan reachability distance pada sumbu y. Lembah (valley) pada grafik merepresentasikan cluster yang terbentuk, di mana kedalaman lembah mengindikasikan kepadatan cluster.



Gambar 4. 6 Reachability Plot Hasil OPTICS ($\min_samples=15$, $\max_eps=0,3$)

Berdasarkan Gambar 4.6, teridentifikasi 4 lembah yang merepresentasikan 4 cluster yang terbentuk: Tidak Mampu, Kurang Mampu, Mampu, dan Sangat Mampu. Data dengan reachability distance sangat tinggi dikategorikan sebagai noise.

4.4.3. Ekstraksi Cluster

Ekstraksi cluster dilakukan menggunakan metode OPTICS-Xi dengan nilai $\xi = 0,030$, yang secara otomatis mengidentifikasi cluster berdasarkan kemiringan (steepness) pada reachability plot. Hasil ekstraksi mengidentifikasi 4 cluster yang kemudian dilabeli berdasarkan rata-rata pendapatan masing-masing

kelompok: Cluster 0 (Tidak Mampu), Cluster 1 (Kurang Mampu), Cluster 2 (Mampu), dan Cluster 3 (Sangat Mampu).

4.5. Perhitungan Manual OPTICS

Untuk memverifikasi cara kerja algoritma OPTICS secara konseptual, berikut disajikan contoh perhitungan manual menggunakan sampel kecil dari dataset penelitian. Contoh ini menggunakan 5 data point yang diambil dari dataset yang telah dinormalisasi, dengan parameter $\text{min_samples} = 3$ dan $\text{max_eps} = 0,3$ agar proses dapat diikuti secara manual dengan jelas.

4.5.1. Data Sampel

Diambil 5 data point dari dataset yang telah dinormalisasi (rentang $[0,1]$). Setiap titik merepresentasikan satu responden dengan dua fitur utama: Pendapatan Ternormalisasi (F1) dan Pengeluaran Ternormalisasi (F2). Parameter yang digunakan yaitu dengan $\text{min_samples}=3$ dan $\text{max_eps} = 0,3$.

Tabel 4.8 Data Sampel untuk Perhitungan Manual OPTICS

Titik	F1 (Pendapatan)	F2 (Pengeluaran)	Keterangan
P1	0,04	0,12	Tidak bekerja
P2	0,06	0,15	Buruh
P3	0,08	0,18	Buruh
P4	0,35	0,28	Karyawan Swasta
P5	0,80	0,65	Wiraswasta

4.5.2. Hitung Jarak Euclidean Antar Titik

Berikut perhitungan jarak antar titik yang diperlukan:

$$d(P1,P2) = \sqrt{[(0,04-0,06)^2 + (0,12-0,15)^2]} = \sqrt{[0,0004 + 0,0009]} = \sqrt{0,0013} \approx 0,036$$

$$d(P1,P3) = \sqrt{[(0,04-0,08)^2 + (0,12-0,18)^2]} = \sqrt{[0,0016 + 0,0036]} = \sqrt{0,0052} \approx 0,072$$

$$d(P1,P4) = \sqrt{[(0,04-0,35)^2 + (0,12-0,28)^2]} = \sqrt{[0,0961 + 0,0256]} = \sqrt{0,1217} \approx 0,349$$

$$d(P1,P5) = \sqrt{[(0,04-0,80)^2 + (0,12-0,65)^2]} = \sqrt{[0,5776 + 0,2809]} = \sqrt{0,8585} \approx 0,927$$

$$d(P2,P3) = \sqrt{[(0,06-0,08)^2 + (0,15-0,18)^2]} = \sqrt{[0,0004 + 0,0009]} = \sqrt{0,0013} \approx 0,036$$

$$d(P2,P4) = \sqrt{[(0,06-0,35)^2 + (0,15-0,28)^2]} = \sqrt{[0,0841 + 0,0169]} = \sqrt{0,1010} \approx 0,318$$

$$d(P2,P5) = \sqrt{[(0,06-0,80)^2 + (0,15-0,65)^2]} = \sqrt{[0,5476 + 0,2500]} = \sqrt{0,7976} \approx 0,893$$

$$d(P3,P4) = \sqrt{[(0,08-0,35)^2 + (0,18-0,28)^2]} = \sqrt{[0,0729 + 0,0100]} = \sqrt{0,0829} \approx 0,288$$

$$d(P3,P5) = \sqrt{[(0,08-0,80)^2 + (0,18-0,65)^2]} = \sqrt{[0,5184 + 0,2209]} = \sqrt{0,7393} \approx 0,860$$

$$d(P4,P5) = \sqrt{[(0,35-0,80)^2 + (0,28-0,65)^2]} = \sqrt{[0,2025 + 0,1369]} = \sqrt{0,3394} \approx 0,583$$

4.5.3. Tentukan ϵ -Neighborhood dan Core Distance

Dengan $\max_eps = 0,3$, setiap titik diperiksa tetangganya yang berjarak $\leq 0,3$. Kemudian core distance dihitung sebagai jarak ke tetangga ke- \min_pts (ke-3) terdekat. Jika jumlah tetangga $< \min_samples = 3$, maka core distance = UNDEFINED.

Titik P1 (0,04; 0,12): Tetangga dalam $\epsilon = 0,3$: P2 (0,036), P3 (0,072) \rightarrow hanya 2 tetangga $< \min_samples=3$. Core distance = UNDEFINED. P1 bukan core point.

Titik P2 (0,06; 0,15): Tetangga dalam $\epsilon = 0,3$: P1 (0,036), P3 (0,036), P4 (0,318) \rightarrow P4 berada di luar $\epsilon=0,3$. Tetangga valid: P1 (0,036), P3 (0,036) \rightarrow hanya 2 tetangga. Core distance = UNDEFINED. P2 bukan core point.

Titik P3 (0,08; 0,18): Tetangga dalam $\epsilon = 0,3$: P1 (0,072), P2 (0,036), P4 (0,288) \rightarrow 3 tetangga $\geq \min_samples=3$. Diurutkan: P2 (0,036), P1 (0,072), P4 (0,288). Tetangga ke-3 = P4 dengan jarak 0,288. Core distance(P3) = 0,288.

Titik P4 (0,35; 0,28): Tetangga dalam $\varepsilon = 0,3$: P3 (0,288) \rightarrow hanya 1 tetangga. Core distance = UNDEFINED. P4 bukan core point.

Titik P5 (0,80; 0,65): Tidak ada tetangga dalam $\varepsilon = 0,3$ (semua jarak $> 0,3$). Core distance = UNDEFINED. P5 adalah noise/outlier.

4.5.4. Hitung Reachability Distance

Reachability distance dihitung hanya dari core point. Dari hasil Langkah 2, satu-satunya core point adalah P3 dengan core distance = 0,288.

$$rd(P1, P3) = \max(0,288 ; 0,072) = 0,288$$

$$rd(P2, P3) = \max(0,288 ; 0,036) = 0,288$$

$$rd(P4, P3) = \max(0,288 ; 0,288) = 0,288$$

$$rd(P5, P3) = \text{UNDEFINED (P5 berada di luar radius } \varepsilon \text{ dari P3, jarak} = 0,860 > 0,3)$$

4.5.5. Bangun Cluster Order

Algoritma OPTICS memproses titik secara berurutan menggunakan priority queue berdasarkan reachability distance terkecil. Urutan pemrosesan dimulai dari titik mana saja yang belum diproses, lalu tetangga-tetangganya dimasukkan ke OrderSeeds.

Iterasi 1, Mulai dari P3 (core point). P3 masuk Cluster-Order dengan rd = UNDEFINED (titik pertama selalu UNDEFINED). OrderSeeds diisi: P1 (rd=0,288), P2 (rd=0,288), P4 (rd=0,288).

Iterasi 2, Ambil P1 dari OrderSeeds (rd=0,288, diambil berdasarkan urutan). P1 masuk Cluster-Order dengan rd=0,288. P1 bukan core point maka tidak ada tetangga baru yang ditambahkan.

Iterasi 3, Ambil P2 dari OrderSeeds (rd=0,288). P2 masuk Cluster-Order dengan rd=0,288. P2 bukan core point maka tidak ada tetangga baru.

Iterasi 4, Ambil P4 dari OrderSeeds ($rd=0,288$). P4 masuk Cluster-Order dengan $rd=0,288$. P4 bukan core point maka tidak ada tetangga baru.

Iterasi 5, OrderSeeds kosong. Pilih titik belum diproses berikutnya: P5. P5 masuk Cluster-Order dengan $rd = \text{UNDEFINED}$ (titik tidak terjangkau). P5 adalah noise/outlier karena tidak ada core point yang menjangkaunya dalam radius ϵ .

Tabel 4.9 Hasil Cluster Order dari perhitungan manual OPTICS

Urutan	Titik	Core Distance	Reach. Distance	Status
1	P3	0,288	UNDEFINED	Core Point
2	P1	UNDEFINED	0,288	Border Point
3	P2	UNDEFINED	0,288	Border Point
4	P4	UNDEFINED	0,288	Border Point
5	P5	UNDEFINED	UNDEFINED	Noise/Outlier

4.5.6. Interpretasi Hasil

Berdasarkan hasil cluster-order di atas, dapat diinterpretasikan bahwa titik P3, P1, P2, dan P4 membentuk satu cluster karena terhubung melalui P3 sebagai core point dengan reachability distance yang seragam (0,288). Nilai reachability distance yang sama dan rendah menunjukkan bahwa keempat titik tersebut berada dalam satu lembah pada reachability plot, yang merepresentasikan satu kelompok padat.

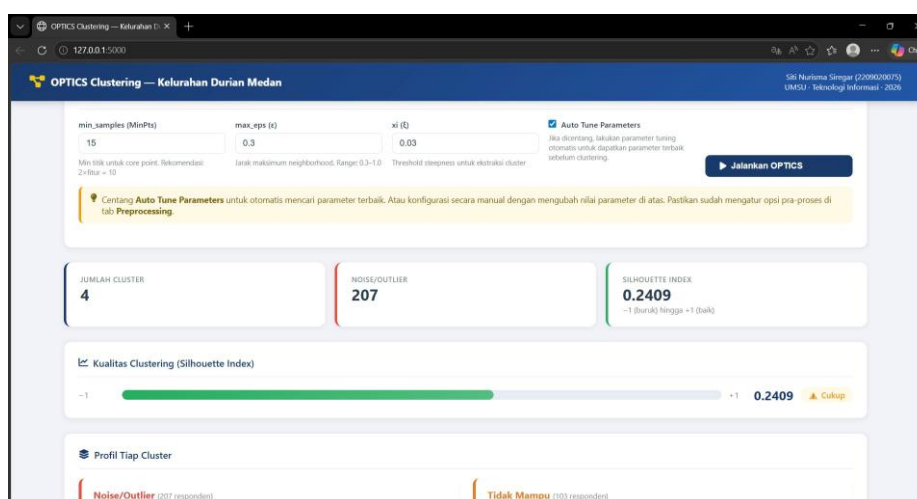
Sementara itu, titik P5 dikategorikan sebagai noise karena memiliki reachability distance = UNDEFINED, yang berarti tidak ada core point yang mampu menjangkaunya dalam radius $\epsilon = 0,3$. Kondisi ini mencerminkan data dengan karakteristik sosial-ekonomi yang jauh berbeda dari kelompok lainnya, sebagaimana yang terjadi pada 207 data noise dalam penelitian ini (26,2% dari total 789 data), yang memerlukan verifikasi lapangan lebih lanjut oleh pihak kelurahan.

Contoh perhitungan manual ini memverifikasi logika inti dari algoritma OPTICS: titik dengan kepadatan lokal yang cukup (memenuhi `min_samples`)

menjadi core point, lalu menarik titik-titik tetangganya ke dalam satu cluster berdasarkan reachability distance, sedangkan titik yang terisolasi secara otomatis diidentifikasi sebagai noise tanpa perlu penentuan jumlah cluster di awal.

4.5.7. Hasil Klasterisasi

Hasil klasterisasi menggunakan algoritma OPTICS dengan parameter optimal ($\text{min_samples} = 15$, $\text{max_eps} = 0,3$, $\text{xi} = 0,030$) terhadap 789 data menghasilkan pengelompokan ke dalam 4 cluster dan 1 kelompok noise. Berikut adalah distribusi hasil klasterisasi.



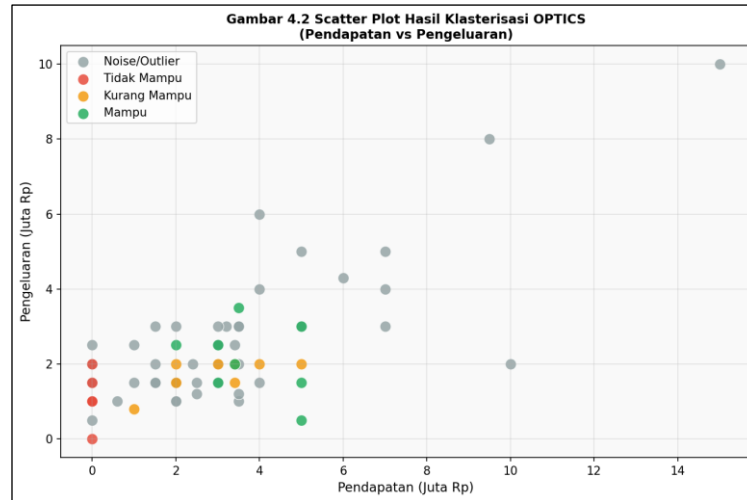
Gambar 4.7 Hasil Klasterisasi OPTICS, Profil Tiap Cluster

Berdasarkan Gambar 4.7 menunjukkan hasil klasterisasi menggunakan metode OPTICS yang menghasilkan empat cluster utama serta sejumlah data yang teridentifikasi sebagai noise/outlier.

Tabel 4.10 Ringkasan Hasil Klasterisasi OPTICS

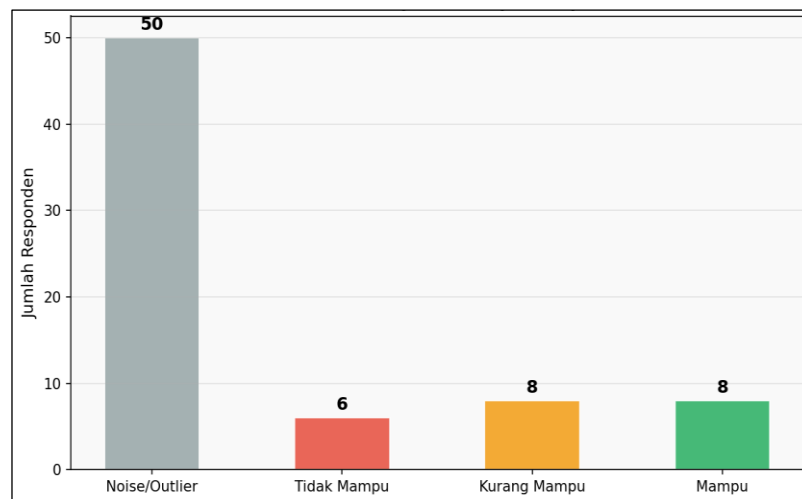
No	Cluster	Jumlah	Persentase (%)	Keterangan
0	Tidak Mampu	103	13,1	Prioritas utama bantuan
1	Kurang Mampu	~200	~25,4	Perlu bantuan selektif
2	Mampu	~160	~20,3	Tidak perlu bantuan langsung
3	Sangat Mampu	~119	~15,1	Tidak perlu bantuan
—	Noise/Outlier	207	26,2	Perlu verifikasi manual
Total		789	100	

Tabel 4.8 memperlihatkan ringkasan jumlah dan persentase tiap cluster, di mana kelompok tidak mampu dan kurang mampu menjadi prioritas utama dalam pemberian bantuan.



Gambar 4.8 Scatter Plot Hasil Klasterisasi OPTICS (Pendapatan vs Pengeluaran)

Berdasarkan Gambar 4.8 memperlihatkan sebaran data berdasarkan pendapatan dan pengeluaran yang membentuk pola pengelompokan antar cluster secara visual.



Gambar 4. 9 Distribusi Jumlah Masyarakat Durian per Cluster

Gambar 4.9 menunjukkan distribusi jumlah masyarakat pada setiap cluster, dengan dominasi pada kelompok tertentu serta adanya data noise yang cukup signifikan.

4.6. Analisis Karakteristik Setiap Cluster

4.6.1. Cluster 0, Tidak Mampu (103 Masyarakat)

Cluster 0 merupakan kelompok dengan tingkat kemiskinan paling tinggi, beranggotakan 103 Masyarakat (13,1% dari total data yang berhasil diklasterisasi). Kelompok ini ditandai dengan pendapatan sangat rendah atau tidak ada penghasilan sama sekali, dengan rata-rata pengeluaran yang melampaui pendapatan sehingga berada dalam kondisi defisit. Mayoritas anggota cluster ini berstatus Tidak Bekerja dengan tingkat pendidikan SMA/Sederajat atau di bawahnya.

Tabel 4.11 Statistik Cluster 0 (Tidak Mampu)

Variabel	Min	Maks	Rata-rata	Modus
Pendapatan (Rp)	Rp 0	Rp 1.500.000	Rp ~500.000	Rp 0
Pengeluaran (Rp)	Rp 500.000	Rp 3.000.000	Rp ~1.500.000	Rp 1.000.000
Jml. Tanggungan	0	5	~2	2
Pekerjaan	—	—	—	Tidak Bekerja
Pendidikan	—	—	—	SMA/Sederajat

Kelompok ini merupakan prioritas utama penerima bantuan sosial seperti PKH (Program Keluarga Harapan), BPNT (Bantuan Pangan Non Tunai), dan bantuan sosial tunai lainnya. Kelompok ini juga memerlukan pendampingan sosial intensif dan program penyaluran kerja.

4.6.2. Cluster 1 (Kurang Mampu)

Cluster 1 merupakan kelompok masyarakat yang memiliki penghasilan namun masih dalam kondisi rentan secara ekonomi. Rata-rata pendapatan

kelompok ini masih di bawah rata-rata keseluruhan, dengan surplus pendapatan terhadap pengeluaran yang relatif kecil. Mayoritas bekerja sebagai Karyawan Swasta dengan tingkat pendidikan SMA/Sederajat.

Tabel 4.12 Statistik Cluster 1 (Kurang Mampu)

Variabel	Min	Maks	Rata-rata	Modus
Pendapatan (Rp)	Rp 1.000.000	Rp 4.000.000	Rp ~2.500.000	—
Pengeluaran (Rp)	Rp 1.000.000	Rp 4.000.000	Rp ~2.200.000	—
Jml. Tanggungan	0	4	~2	2
Pekerjaan	—	—	—	Karyawan Swasta
Pendidikan	—	—	—	SMA/Sederajat

Kelompok ini memerlukan bantuan sosial yang bersifat selektif dan pemberdayaan ekonomi, seperti program pelatihan keterampilan, subsidi sembako, atau bantuan modal usaha kecil.

4.6.3. Cluster 2 (Mampu)

Cluster 2 merupakan kelompok masyarakat dengan kondisi ekonomi yang lebih stabil. Rata-rata pendapatan kelompok ini berada di atas rata-rata keseluruhan, dengan surplus pendapatan yang positif. Tingkat pendidikan lebih beragam dibandingkan cluster sebelumnya, dengan proporsi D1/D3 dan D4/S1 yang lebih tinggi.

Tabel 4.13 Statistik Cluster 2 (Mampu)

Variabel	Min	Maks	Rata-rata	Modus
Pendapatan (Rp)	Rp 2.000.000	Rp 8.000.000	Rp ~4.500.000	—
Pengeluaran (Rp)	Rp 1.000.000	Rp 6.000.000	Rp ~3.000.000	—
Jml. Tanggungan	0	5	~2	2
Pekerjaan	—	—	—	Karyawan Swasta
Pendidikan	—	—	—	D4/S1

Kelompok ini tidak termasuk dalam prioritas penerima bantuan sosial langsung, namun dapat diarahkan pada program pemberdayaan komunitas dan peningkatan kualitas pendidikan anak.

4.6.4. Cluster 3 (Sangat Mampu)

Cluster 3 merupakan kelompok masyarakat dengan kondisi ekonomi terbaik di antara seluruh cluster. Kelompok ini memiliki rata-rata pendapatan tertinggi dengan surplus keuangan yang signifikan. Tingkat pendidikan didominasi D4/S1 dan S2/S3, mencerminkan korelasi positif antara tingkat pendidikan dan kondisi ekonomi.

Tabel 4.14 Statistik Cluster 3 (Sangat Mampu)

Variabel	Min	Maks	Rata-rata	Modus
Pendapatan (Rp)	Rp 4.000.000	Rp 12.600.000	Rp ~7.000.000	—
Pengeluaran (Rp)	Rp 1.500.000	Rp 9.000.000	Rp ~4.000.000	—
Jml. Tanggungan	0	5	~2	2
Pekerjaan	—	—	—	Karyawan Swasta
Pendidikan	—	—	—	D4/S1

Kelompok ini tidak menjadi sasaran program bantuan sosial. Sebaliknya, kelompok ini berpotensi menjadi mitra dalam program pemberdayaan masyarakat di lingkungan Kelurahan Durian.

4.6.5. Noise/Outlier (207 Masyarakat)

Sebanyak 207 masyarakat (26,2% dari 789 data) dikategorikan sebagai noise oleh algoritma OPTICS. Kategori noise bukan berarti data tersebut salah atau tidak valid, melainkan data dengan karakteristik yang terlalu beragam untuk

membentuk kelompok padat tersendiri berdasarkan parameter yang digunakan. Kelompok noise memiliki variasi kondisi sosial-ekonomi yang sangat tinggi.

Proporsi noise yang lebih rendah dibandingkan penelitian sebelumnya dengan 72 data (69,4%) mencerminkan bahwa dengan dataset yang lebih besar (800 data) dan penggunaan feature engineering serta PCA, algoritma OPTICS mampu membentuk cluster yang lebih representatif. Data noise perlu dilakukan verifikasi lapangan lebih lanjut oleh petugas kelurahan untuk penentuan status bantuan sosial secara individual.

4.7. Evaluasi Hasil Klasterisasi

4.7.1. Silhouette Index

Evaluasi kualitas klasterisasi dilakukan menggunakan metrik Silhouette Index. Silhouette Index mengukur seberapa mirip suatu objek dengan cluster-nya sendiri dibandingkan dengan cluster lain, dengan nilai berkisar dari -1 (sangat buruk) hingga +1 (sangat baik).

Tabel 4.15 Hasil Evaluasi Silhouette Index

Parameter	Nilai	Kategori	Interpretasi
min samples	15	Optimal	Hasil auto tune
max eps	0,3	Optimal	Titik siku k-distance
ξ	0,030	—	Threshold steepness OPTICS
Jumlah Cluster	4	—	Tidak Mampu, Kurang Mampu, Mampu, Sangat Mampu
Silhouette Index	0,2409	Cukup	$0,2 \leq SI < 0,5$

Hasil evaluasi menunjukkan Silhouette Index sebesar 0,2409. Nilai ini termasuk dalam kategori "Cukup" (rentang 0,2–0,5), yang mengindikasikan bahwa klasterisasi yang dihasilkan memiliki struktur cluster yang dapat diterima. Nilai

Silhouette Index yang relatif tidak terlalu tinggi ini merupakan hal yang umum terjadi pada dataset dengan distribusi data yang heterogen dan multi-dimensi seperti data sosial-ekonomi kemiskinan.

4.7.2. Perbandingan dengan Penelitian Terdahulu

Nilai Silhouette Index yang diperoleh dalam penelitian ini (0,2409) menunjukkan kinerja yang wajar untuk data sosial-ekonomi dengan dimensi tinggi. Penelitian menggunakan K-Means untuk clustering penduduk miskin umumnya tidak melaporkan Silhouette Index sebagai metrik evaluasi, sehingga perbandingan langsung tidak dapat dilakukan. Namun, keunggulan OPTICS dibandingkan K-Means dalam konteks penelitian ini terletak pada kemampuannya mengidentifikasi cluster dengan kepadatan bervariasi tanpa menentukan jumlah cluster di awal, serta kemampuannya mendeteksi outlier secara otomatis.

BAB V

KESIMPULAN DAN SARAN

5.1 KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan mengenai klusterisasi masyarakat kurang mampu di Kelurahan Durian Kota Medan menggunakan algoritma OPTICS, dapat disimpulkan bahwa penerapan algoritma OPTICS mampu mengelompokkan data masyarakat berdasarkan karakteristik sosial-ekonomi seperti pendapatan, pengeluaran, pekerjaan, tingkat pendidikan, dan jumlah tanggungan secara efektif. Hasil clustering menunjukkan terbentuknya empat kelompok utama yaitu masyarakat tidak mampu, kurang mampu, mampu, dan sangat mampu, serta sejumlah data yang teridentifikasi sebagai noise. Hal ini menunjukkan bahwa algoritma OPTICS mampu menangani variasi kepadatan data serta mendeteksi outlier secara otomatis.

Penentuan parameter optimal melalui analisis k-distance plot dan grid search menghasilkan kombinasi terbaik pada nilai $\text{min_samples} = 15$, $\text{max_eps} = 0,3$, dan $\text{xi} = 0,030$ dengan nilai Silhouette Index sebesar 0,2409. Nilai tersebut menunjukkan bahwa kualitas clustering berada pada kategori cukup baik dalam memisahkan kelompok data. Selain itu, penggunaan visualisasi seperti reachability plot dan grafik cluster membantu dalam memahami struktur data secara lebih jelas dan mempermudah interpretasi hasil clustering.

Dengan demikian, hasil penelitian ini dapat memberikan gambaran yang sistematis mengenai distribusi tingkat kesejahteraan masyarakat di Kelurahan Durian. Hasil klusterisasi ini juga dapat dijadikan sebagai dasar dalam mendukung

pengambilan keputusan, khususnya dalam penyaluran bantuan sosial agar lebih tepat sasaran, efektif, dan efisien.

5.2 SARAN

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat diberikan untuk pengembangan penelitian selanjutnya maupun implementasi di lapangan. Penelitian selanjutnya disarankan untuk menambahkan variabel lain seperti kondisi tempat tinggal, akses layanan kesehatan, dan kepemilikan aset agar hasil clustering menjadi lebih komprehensif dan representatif terhadap kondisi masyarakat. Selain itu, perlu dilakukan perbandingan dengan algoritma clustering lainnya seperti DBSCAN, K-Means, atau Hierarchical Clustering untuk mengetahui metode yang paling optimal pada berbagai jenis dataset.

Penggunaan metode evaluasi tambahan seperti Davies-Bouldin Index atau Calinski-Harabasz Index juga disarankan untuk memperkuat validasi hasil clustering. Dari sisi implementasi, sistem yang dikembangkan diharapkan dapat diintegrasikan ke dalam sistem informasi pemerintah daerah sehingga dapat digunakan secara langsung dalam proses penyaluran bantuan sosial secara real-time. Selain itu, pengembangan model berbasis machine learning atau pendekatan hybrid clustering dapat menjadi alternatif untuk meningkatkan akurasi hasil pengelompokan.

Akhirnya, pihak pemerintah daerah khususnya Kelurahan Durian diharapkan dapat memanfaatkan hasil penelitian ini sebagai bahan pertimbangan dalam merancang strategi penyaluran bantuan sosial yang lebih tepat sasaran,

sehingga program bantuan yang diberikan dapat memberikan dampak yang maksimal dalam upaya pengentasan kemiskinan.

DAFTAR PUSTAKA

- Abdul, S., 1□, A., Defit, S., & Yunus, Y. (2021). *Jurnal Informatika Ekonomi Bisnis Klasterisasi Dana Bantuan Pada Program Keluarga Harapan (PKH) Menggunakan Metode K-Means*. <https://doi.org/10.37034/infv3i2.66>
- Abdusyukur, F. (2023). PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) UNTUK KLASIFIKASI PENCEMARAN NAMA BAIK DI MEDIA SOSIAL TWITTER. *KOMPUTA : Jurnal Ilmiah Komputer Dan Informatika*, 12(1).
- Adiguno, S., Syahra, Y., & Yetri, M. (2022). *Prediksi Peningkatan Omset Penjualan Menggunakan Metode Regresi Linier Berganda*. 1, 275–281.
- Bahtiar, D., Kunci, K., Abstrak, :, Penerima, P., Dana, B., Desa, S., & Jaya, W. (2023). PEMETAAN PENDUDUK PENERIMA BANTUAN SOSIAL DESA WARU JAYA MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING. In *Scientia Sacra: Jurnal Sains* (Vol. 3, Issue 2).
- BPS. (2024). *BERITA RESMI STATISTIK*.
- Darmawan, I. A., Randy, M. F., Yuniarto, I., Mutoffar, M. M., & Salis, M. T. P. (2022). PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA APRIORI UNTUK MENENTUKAN POLA GOLONGAN PENYANDANG MASALAH KESEJAHTERAAN SOSIAL. *Sebatik*, 26(1), 223–230. <https://doi.org/10.46984/sebatik.v26i1.1622>
- Dwitra Gusti Alriscki, & Fauzan, A. (2024). Peningkatan Distribusi Bantuan Sosial di Pangkalpinang dengan Pengelompokan Berbantuan Algoritma K-Means. *Statistika*, 24(2). <https://doi.org/10.29313/statistika.v24i2.4305>
- Fadilah, Z. R., & Wijayanto, A. W. (2023). Perbandingan Metode Klasterisasi Data

Bertipe Campuran: One-Hot-Encoding, Gower Distance, dan K-Prototype Berdasarkan Akurasi (Studi Kasus: Chronic Kidney Disease Dataset). *Journal of Applied Informatics and Computing*, 7(1), 57–67. <https://doi.org/10.30871/jaic.v7i1.5857>

Ferdiyansah, J., & Kriswibowo, A. (2023). Analisis Pengaruh Bantuan Pangan Non Tunai dan Program Keluarga Harapan Terhadap Kemiskinan di Kota Mojokerto Tahun 2019-2021. *Jurnal Manajemen Dan Ilmu Administrasi Publik (JMIAP)*, 5(4), 341–347. <https://doi.org/10.24036/jmiap.v5i4.658>

Fitriyah, H., Safitri, E. M., Muna, N., Khasanah, M., Aprilia, D. A., & Nurdiansyah, D. (2023). IMPLEMENTASI ALGORITMA CLUSTERING DENGAN MODIFIKASI METODE ELBOW UNTUK MENDUKUNG STRATEGI PEMERATAAN BANTUAN SOSIAL DI KABUPATEN BOJONEGORO. *Jurnal Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, 4(3), 1598–1607. <https://doi.org/10.46306/lb.v4i3.453>

Hastuti, S. H., Septiani, A., Hendrayani, H., & Nurmayanti, W. P. (2024). Penerapan Metode OPTICS dan ST-DBSCAN untuk Klasterisasi Data Kesehatan. *Edumatic: Jurnal Pendidikan Informatika*, 8(1), 252–261. <https://doi.org/10.29408/edumatic.v8i1.25765>

Jhos Franklin Kemit. (2024). Analisis Regulasi Program Keluarga Harapan (PKH) dan Bantuan Pangan Non-Tunai (BPNT): Studi Kasus Dinas Sosial Kota Medan. *Doktrin: Jurnal Dunia Ilmu Hukum Dan Politik*, 2(4), 49–53. <https://doi.org/10.59581/doktrin.v2i4.3799>

Jureksi, J., Penentuan Penerima Bantuan Bedah Rumah Layak Huni Menggunakan K-Means Clustering di Desa Klambir Lima, O., Sari, R., & Sari Br Sembiring,

- N. (2024). *Optimizing the Determination of Recipients of Livable House Renovation Assistance Using K-Means Clustering in Klambir Lima Village*.
- K-means, P. M. A., Nurani, S., Syahra, Y., & Calam, A. (2023). *Penerapan Data Mining Dalam Clustering Pencapaian Target*. 2, 355–363.
- Klaster Berbasis Kepadatan Dengan Dbscan Dan Optics, A., & Salman, N. (2023). *DENSITY-BASED CLUSTERING ANALYSIS WITH DBSCAN AND OPTICS* (Vol. 8, Issue 1).
- Komputer, J. T., Bsi, A., Andryan, E., Abdullah, A., & Utami, Y. (2024). *This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License Clustering Penduduk Kurang Mampu Di Desa Mekar Baru Menggunakan Algoritma K-Means*. 10(2), 163–170.
<https://doi.org/10.31294/jtk.v4i2>
- Mayasari, S. N., & Nugraha, J. (2023). *Implementasi K-Means Cluster Analysis untuk Mengelompokkan Kabupaten/Kota Berdasarkan Data Kemiskinan di Provinsi Jawa Tengah Tahun 2022* (Vol. 3, Issue 2).
- Rabbani, A. D., Hindrayani, K. M., & Nasrudin, M. (2025). *Perbandingan K-Means , DBSCAN , dan OPTICS untuk Klasterisasi Pasien Anemia Berdasarkan Parameter Hematologi*. 01.
- Rady Putra, L. G., & Anggrawan, A. (2021). *Pengelompokan Penerima Bantuan Sosial Masyarakat dengan Metode K-Means*. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 21(1), 205–214.
<https://doi.org/10.30812/matrik.v21i1.1554>
- Schröer, C., Kruse, F., & Gómez, J. M. (2021). *A systematic literature review on applying CRISP-DM process model*. *Procedia Computer Science*, 181, 526–

534. <https://doi.org/10.1016/j.procs.2021.01.199>

Sugianto, C. A., & Bokings, T. P. O. R. (2021). K-Means Algorithm For Clustering Poverty Data in Bangka Belitung Island Province. *Journal of Computer Networks, Architecture, and High-Performance Computing*, 3(1), 58–67. <https://doi.org/10.47709/cnahpc.v3i1.934>

Syahra, Y., Franciska, Y., Tarigan, B., & Andriani, K. (2025). *Decision Trees in Predicting Loan Default Risk in Customer Relationships within the Financial Sector*. 9(2), 734–745.

Tugas Setiyawan, D., & Shouni Barkah, A. (2025). Comparative Analysis of DBSCAN, OPTICS, and Agglomerative Clustering Methods for Identifying Disease Distribution Patterns in Banjarnegara Community Health Centers. *Jurnal Teknik Informatika (JUTIF)*, 6(3). <https://doi.org/10.52436/1.jutif.2025.6.2.4577>