

Pemetaan Karakteristik Wilayah Miskin di Jawa Timur Menggunakan Spectral Clustering untuk Penentuan Target Kebijakan

Rahmalia Anindya Herdianti¹, Wahyu Syaifulah Jauharis Saputra^{*2}, Dwi Arman Prasetya³

^{1,2,3} Sains Data, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Indonesia.

Email: ¹rahmaliaanindya123@gmail.com, ²wahyu.s.j.saputra.if@upnjatim.ac.id,

³arman.prasetya.sada@upnjatim.ac.id

Abstrak

Tingkat kemiskinan yang masih tinggi di Jawa Timur merupakan permasalahan serius yang memerlukan perhatian dan penanganan berbasis data yang komprehensif. Ketimpangan sosial ekonomi antarwilayah menyebabkan sebagian daerah tertinggal dalam pembangunan, sehingga pengelompokan wilayah miskin menjadi langkah strategis untuk membantu pemerintah dalam menentukan prioritas serta arah kebijakan pengentasan kemiskinan yang lebih efektif. Penelitian ini bertujuan untuk memetakan karakteristik wilayah miskin di Jawa Timur menggunakan algoritma Spectral Clustering dengan 11 indikator sosial ekonomi sebagai variabel analisis. Berdasarkan hasil pembobotan kontribusi variabel terhadap pembentukan klaster, indikator yang paling berpengaruh adalah garis kemiskinan (0,259418), rata-rata lama sekolah (0,259067), harapan lama sekolah (0,147613), dan indeks pembangunan manusia (0,133441). Hasil pengelompokan menunjukkan adanya dua klaster utama, yaitu wilayah dengan tingkat kemiskinan tinggi (26 kabupaten/kota) dan wilayah dengan tingkat kemiskinan rendah (12 kabupaten/kota). Evaluasi kualitas model menghasilkan nilai Davies-Bouldin Index (0,4401) dan Silhouette Score (0,6655), yang menunjukkan bahwa metode ini mampu membentuk kelompok dengan pemisahan yang cukup baik. Temuan ini diharapkan dapat menjadi dasar dalam perumusan kebijakan intervensi yang lebih terarah, berkelanjutan, dan berkeadilan sosial di Jawa Timur.

Kata kunci: Kemiskinan, Spectral Clustering, Analisis Klaster, Jawa Timur, Pemetaan Wilayah.

Abstract

The persistently high poverty rate in East Java is a serious problem that requires comprehensive data-based attention and management. Socioeconomic disparities between regions have left some regions behind in development, thus categorizing poor regions is a strategic step to assist the government in determining priorities and the direction of more effective poverty alleviation policies. This study aims to map the characteristics of poor regions in East Java using the Spectral Clustering algorithm with 11 socioeconomic indicators as analysis variables. Based on the results of the weighting of variable contributions to cluster formation, the most influential indicators are the poverty line (0.259418), average years of schooling (0.259067), expected years of schooling (0.147613), and the human development index (0.133441). The clustering results indicate the existence of two main clusters, namely areas with high poverty rates (26 districts/cities) and areas with low poverty rates (12 districts/cities). The evaluation of model quality yielded a Davies-Bouldin Index (0.4401) and Silhouette Score (0.6655), indicating that this method is capable of forming clusters with fairly good separation. These findings are expected to form the basis for formulating more targeted, sustainable, and socially just intervention policies in East Java.

Keywords: Poverty, Spectral Clustering, Cluster Analysis, East Java, Regional Mapping.

This work is an open access article and licensed under a Creative Commons Attribution-NonCommercial ShareAlike 4.0 International (CC BY-NC-SA 4.0)



1. PENDAHULUAN

Indonesia merupakan salah satu negara berkembang, dimana terdapat banyak masalah yang terjadi. Salah satu masalah tersebut adalah kemiskinan. Kemiskinan merupakan suatu kondisi dimana seseorang tidak dapat memenuhi kebutuhan untuk menjalani kehidupan, baik berupa aspek pangan, sandang, pendidikan, maupun papan. Kemiskinan dianggap sebagai masalah yang penting karena berkaitan dengan hajat hidup orang banyak [1]. Berdasarkan data Badan Pusat Statistik (BPS), pada Maret 2023 jumlah penduduk miskin di Indonesia mencapai 25,9 juta jiwa atau sekitar 9,36% dari total populasi.

Sementara itu, Jawa Timur merupakan provinsi dengan jumlah penduduk terbesar kedua di Indonesia dan menjadi provinsi termiskin ketiga di Pulau Jawa dengan persentase penduduk miskin mencapai 10,35%. Kondisi ini menunjukkan masih adanya disparitas kesejahteraan antarwilayah di provinsi tersebut. Jawa Timur sering mengalami variasi tingkat kemiskinan yang menjadi penyebab hasil pembangunan tidak merata. Selain itu, terdapat perbedaan standar hidup minimum di berbagai daerah [2]. Salah satu cara untuk memperoleh wawasan dari data yang belum memiliki label adalah dengan melakukan pengelompokan atau klasterisasi terhadap data indikator [3]. Melalui pendekatan ini, dapat diidentifikasi perbedaan tingkat kemiskinan antar wilayah di Provinsi Jawa Timur serta berbagai faktor yang memengaruhi kondisi tersebut.

Pengelompokan ini memungkinkan pemerintah untuk memprioritaskan daerah dengan tingkat kemiskinan yang tinggi sehingga sumber daya dapat dialokasikan secara optimal. Selain itu, dengan memahami karakteristik masing-masing kelompok, pemerintah dapat merancang program pemberdayaan yang lebih relevan dan sesuai dengan kondisi spesifik di masing-masing daerah. Penelitian terdahulu telah memanfaatkan algoritma Spectral Clustering dalam menganalisis tingkat persebaran vaksinasi COVID-19 di Indonesia berdasarkan data per kategori masyarakat. Pengelompokan dilakukan berdasarkan provinsi dan menghasilkan tiga klaster utama, yaitu provinsi dengan tingkat vaksinasi tinggi, sedang, dan rendah. Evaluasi hasil klaster dilakukan menggunakan Davies-Bouldin Index (DBI) dengan nilai rata-rata sebesar 1,01422, yang menunjukkan bahwa metode ini cukup efektif dalam mengelompokkan wilayah berdasarkan distribusi vaksinasi [4]. Selain itu, penelitian lain membahas topik serupa dengan menggunakan metode K-Medoids untuk mengelompokkan kabupaten/kota di Jawa Timur berdasarkan indikator kemiskinan tahun 2020. Hasilnya menunjukkan dua klaster, di mana klaster pertama didominasi oleh wilayah dengan sanitasi layak, angka harapan hidup, dan tingkat melek huruf yang tinggi, sedangkan klaster kedua menunjukkan karakteristik kemiskinan yang lebih tinggi, seperti tingginya penerima raskin dan persentase penduduk miskin [5]. Sampai saat ini, belum ada penelitian yang secara khusus mengkaji pengelompokan tingkat kemiskinan kabupaten/kota di Jawa Timur dengan menggunakan algoritma Spectral Clustering. Pemilihan metode ini didasarkan pada kemampuannya dalam menangani data dengan struktur kompleks dan non-linear, serta kemampuannya mengidentifikasi pola keterhubungan antarwilayah yang tidak dapat dijangkau oleh metode berbasis jarak seperti K-Medoids atau K-Means [4].

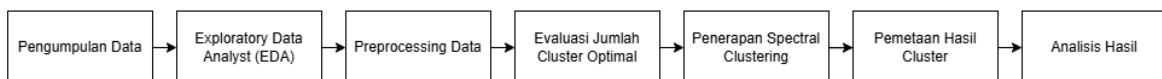
Salah satu algoritma pengelompokan (clustering) yang banyak digunakan dalam Data Mining adalah Spectral Clustering. Algoritma ini berlandaskan pada teori graf dan mengadopsi pendekatan koneksi dalam pengelompokan data. Dalam Spectral Clustering, titik-titik data yang memiliki hubungan atau keterkaitan yang kuat, baik itu secara langsung terhubung atau berbatasan, akan digabungkan ke dalam satu klaster. Hal ini membuat Spectral Clustering efektif dalam menangani data yang tidak memiliki bentuk atau struktur yang jelas [4]. Proses kerja Spectral Clustering dimulai dengan membangun sebuah matriks kemiripan yang menggambarkan kedekatan antar titik data. Matriks ini biasanya dihitung menggunakan fungsi Gaussian, yang bertujuan untuk mengurangi jarak antar titik data sehingga lebih mudah untuk mengidentifikasi kelompok yang serupa. Fungsi Gaussian ini digunakan untuk menilai kemiripan antar pasangan data, di mana titik-titik yang memiliki nilai kemiripan tinggi akan dikelompokkan bersama [6].

Metode ini banyak dipilih dalam konteks pengelompokan data karena kemampuannya dalam menangani data dengan kompleksitas yang tinggi, serta kemampuannya untuk mengidentifikasi klaster non-linear. Oleh karena itu, tujuan utama dari penelitian ini adalah untuk mengelompokkan kabupaten/kota di Jawa Timur berdasarkan tingkat kemiskinan dengan menggunakan Spectral Clustering, untuk mendapatkan pemahaman yang lebih mendalam mengenai distribusi tingkat kemiskinan di provinsi tersebut.

2. METODE PENELITIAN

Metode penelitian merupakan rangkaian langkah yang disusun secara terstruktur dan rasional untuk mencapai tujuan yang telah ditetapkan dalam suatu penelitian. Pada penelitian ini, tahapan awal dimulai dengan pengumpulan data yang mencakup variabel kemiskinan di kabupaten/kota Jawa Timur. Data yang terkumpul dianalisis melalui Exploratory Data Analysis (EDA) untuk mendapatkan gambaran awal mengenai distribusi dan korelasi antar variabel. Selanjutnya, dilakukan preprocessing data yang mencakup penanganan nilai yang hilang dan normalisasi fitur. Untuk menentukan jumlah klaster yang

optimal, digunakan evaluasi dengan Silhouette Score dan Davies-Bouldin Index. Tahap berikutnya adalah penerapan Spectral Clustering untuk pengelompokan data, yang kemudian diikuti dengan pemetaan hasil clustering dan analisis mendalam terhadap hasil yang diperoleh untuk memahami pola kemiskinan di daerah tersebut.



Gambar 1. Alur Penelitian

2.1. Pengumpulan Data

Data penelitian ini berasal dari situs resmi Badan Pusat Statistik (BPS), mencakup 38 kabupaten/kota di Jawa Timur dengan 11 variabel indikator kemiskinan yang merepresentasikan aspek sosial dan ekonomi. Setiap variabel ini diukur dalam bentuk numerik (integer) dan memiliki peran penting dalam memberikan gambaran menyeluruh terhadap kondisi kemiskinan di masing-masing wilayah. Pendekatan ini selaras dengan praktik dalam studi ekonomi pembangunan yang mengukur tingkat kemiskinan melalui indikator garis kemiskinan, sebagaimana diterapkan oleh Badan Pusat Statistik (BPS) [7].

Pemilihan 11 indikator kemiskinan didasarkan pada indikator sosial ekonomi yang digunakan oleh Badan Pusat Statistik (BPS) dan studi sebelumnya terkait analisis kemiskinan regional [7][17]. Indikator tersebut dianggap mampu merepresentasikan dimensi kesejahteraan masyarakat secara komprehensif, meliputi aspek pendidikan, kesehatan, serta kondisi ekonomi rumah tangga. Pendekatan ini juga sejalan dengan penelitian-penelitian terdahulu yang mengukur kemiskinan melalui kombinasi variabel struktural dan nonstruktural untuk memperoleh gambaran multidimensi mengenai tingkat kesejahteraan wilayah [18].

2.2. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) merupakan proses analisis dan visualisasi data untuk memperoleh pemahaman mendalam terhadap informasi yang terkandung di dalamnya. EDA umumnya menggunakan visualisasi grafis, didukung teknik kuantitatif, guna mengeksplorasi data secara terbuka dan memperkuat hasil analisis [8].

Proses ini mencakup pemeriksaan struktur data, identifikasi nilai hilang, serta analisis statistik deskriptif seperti nilai rata-rata, median, dan standar deviasi. Selain itu, EDA melibatkan visualisasi untuk melihat distribusi dan hubungan antar variabel. Tahapan ini sangat penting untuk memastikan kualitas data sebelum masuk ke tahap pemodelan lebih lanjut.

2.3. Preprocessing Data

Preprocessing merupakan proses awal dalam Data mining, dimana data yang berupa data mentah akan diubah menjadi rangkaian unit bahasa yang lebih jelas [9]. Pada tahap preprocessing mencakup penanganan data hilang, duplikasi, normalisasi, serta analisis deskriptif dan korelasi untuk memahami struktur dan hubungan antar data. Data yang teridentifikasi mengalami permasalahan akan dihapus dan tidak disertakan dalam kegiatan penelitian. Langkah ini bertujuan untuk memastikan bahwa data yang digunakan dalam proses analisis memiliki kualitas yang memadai dan dapat dipercaya [10].

2.4. Evaluasi Jumlah Cluster Optimal

Evaluasi kinerja clustering dilakukan menggunakan dua metrik, yaitu Silhouette Score dan Davies-Bouldin Index (DBI). Silhouette Score mengukur seberapa baik objek cocok dengan klusternya dibandingkan dengan klaster lain [11], dengan nilai berkisar antara -1 hingga 1. Nilai yang mendekati 1 menunjukkan pemisahan klaster yang baik dan kohesi yang tinggi dalam klaster [12]. Skor dihitung menggunakan rumus persamaan (1):

$$S_i = \frac{b(o) - a(o)}{\max(a(o), b(o))} \quad (1)$$

Pada persamaan (1) $a(o)$ adalah rata-rata jarak objek terhadap anggota dalam klaster yang sama, dan $b(o)$ adalah rata-rata jarak terhadap objek dalam klaster terdekat.

Sementara itu, Davies-Bouldin Index (DBI) menilai kualitas clustering berdasarkan rasio antara penyebaran dalam klaster dan jarak antar klaster [13]. Nilai DBI yang lebih rendah menunjukkan clustering yang lebih optimal [1]. Pada persamaan (2) merupakan rumus DBI yang akan digunakan untuk melakukan evaluasi jumlah klaster optimal:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} (R_{i,j}) \quad (2)$$

Dengan,

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (3)$$

Dimana pada persamaan (3) SSW mengukur kohesi dalam klaster, dan SSB mengukur separasi antar klaster.

2.5. Penerapan Spectral Clustering

Spectral Clustering merupakan algoritma klasterisasi berbasis graf yang tidak mengasumsikan bentuk tertentu pada struktur klaster [14]. Setiap titik data direpresentasikan sebagai simpul (node) dalam graf, kemudian dilakukan pemisahan berdasarkan pendekatan *graph partitioning* [15]. Setelah proses pemisahan graf dilakukan, setiap objek hasil segmentasi perlu dipisahkan secara individual sebelum melanjutkan ke tahap analisis berikutnya guna memperoleh informasi yang lebih mendalam [16].

Langkah pertama pada persamaan (4) adalah membentuk matriks kemiripan antar data menggunakan fungsi Gaussian kernel:

$$s(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (4)$$

Dengan σ sebagai parameter yang mengontrol skala jarak antar titik.

Selanjutnya, pada persamaan (5) dibentuk matriks Laplacian tak ternormalisasi $L = D - W$, lalu dilakukan normalisasi untuk memperoleh normalized Laplacian:

$$L_{sym} = I - D^{-\frac{1}{2}}WD^{-\frac{1}{2}} \quad (5)$$

Dengan I sebagai matriks identitas. Kemudian, dilakukan dekomposisi eigen terhadap L_{sym} untuk memperoleh k vector eigen terkecil pada persamaan (6) yang digunakan sebagai representasi fitur baru:

$$Z = [v_1, v_2, \dots, v_k] \quad (6)$$

Algoritma K-Means diterapkan pada ruang vektor eigen tersebut untuk memperoleh partisi klaster yang optimal. Serta melakukan evaluasi kualitas pengelompokan menggunakan metrik seperti Silhouette Score dan Davies-Bouldin Index.

2.6. Pemetaan Hasil Cluster

Pemetaan hasil klaster bertujuan untuk mengidentifikasi kabupaten/kota yang termasuk ke dalam masing-masing klaster berdasarkan hasil pengelompokan. Proses ini dilakukan dengan mengaitkan label klaster yang diperoleh dari algoritma dengan data wilayah, sehingga diperoleh informasi mengenai distribusi kabupaten/kota pada setiap klaster.

2.7. Analisis Hasil

Klasifikasi hasil klasterisasi menjadi kategori “tingkat kemiskinan tinggi” dan “tingkat kemiskinan rendah” didasarkan pada nilai rata-rata indikator ekonomi utama dalam setiap klaster. Klaster dengan

nilai rata-rata garis kemiskinan dan persentase penduduk miskin yang lebih tinggi, disertai capaian pendidikan dan indeks pembangunan manusia yang relatif rendah, dikategorikan sebagai wilayah dengan tingkat kemiskinan tinggi. Sebaliknya, klaster dengan karakteristik sebaliknya dikategorikan sebagai wilayah dengan tingkat kemiskinan rendah.

Analisis dilakukan dengan mengkaji karakteristik masing-masing klaster berdasarkan variabel-variabel yang digunakan dalam penelitian. Selanjutnya, dilakukan penelaahan terhadap persebaran wilayah kabupaten/kota pada tiap klaster untuk mengidentifikasi pola atau tren tertentu. Temuan dari analisis ini menjadi landasan dalam merumuskan kesimpulan akhir penelitian.

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

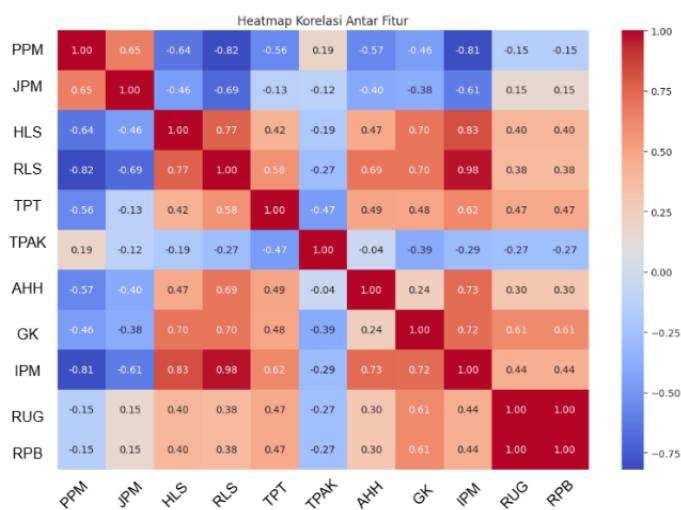
Data yang digunakan dalam penelitian ini berasal dari situs resmi Badan Pusat Statistik (BPS), dengan penekanan pada indikator-indikator yang berkaitan dengan aspek kemiskinan. Dataset mencakup 38 kabupaten/kota di Provinsi Jawa Timur dan terdiri dari 11 variabel.

Tabel 1. Variabel Indikator Kemiskinan

| No. | Variabel | Singkatan Variabel | Tipe |
|-----|---|--------------------|---------|
| 1 | Percentase Penduduk Miskin (%) | PPM | Integer |
| 2 | Jumlah Penduduk Miskin (ribu jiwa) | JPM | Integer |
| 3 | Harapan Lama Sekolah (Tahun) | HLS | Integer |
| 4 | Rata-Rata Lama Sekolah (Tahun) | RLS | Integer |
| 5 | Tingkat Pengangguran Terbuka (%) | TPT | Integer |
| 6 | Tingkat Partisipasi Angkatan Kerja (%) | TPAK | Integer |
| 7 | Angka Harapan Hidup (Tahun) | AHH | Integer |
| 8 | Garis Kemiskinan (Rupiah/Bulan/Kapita) | GK | Integer |
| 9 | Indeks Pembangunan Manusia | IPM | Integer |
| 10 | Rata-rata Upah/Gaji Bersih Pekerja Informal Berdasarkan Lapangan Pekerjaan Utama (Rp) | RUG | Integer |
| 11 | Rata-rata Pendapatan Bersih Sebulan Pekerja Informal berdasarkan Pendidikan Tertinggi - Jumlah (Rp) | RPB | Integer |

3.2. Exploratory Data Analysis (EDA)

Eksplorasi awal terhadap data dilakukan guna memahami struktur serta karakteristik variabel yang digunakan. Hasil pemeriksaan menunjukkan bahwa seluruh variabel berformat numerik dan tidak ditemukan nilai yang hilang. Analisis statistik deskriptif terhadap 11 indikator kemiskinan mengungkapkan adanya perbedaan signifikan antar kabupaten/kota di Jawa Timur. Sebagai contoh, persentase penduduk miskin bervariasi dengan nilai minimum sebesar 3.790000, maksimum 21.610000, dan rata-rata sebesar 10.325789.



Gambar 2. Korelasi Indikator Kemiskinan

Berdasarkan hasil visualisasi heatmap korelasi pada gambar 2, terlihat bahwa Persentase Penduduk Miskin memiliki korelasi negatif yang kuat dengan Indeks Pembangunan Manusia dan Harapan Lama Sekolah. Hal ini menunjukkan bahwa semakin tinggi tingkat pendidikan dan pembangunan manusia, maka tingkat kemiskinan cenderung menurun. Selain itu, Indeks Pembangunan Manusia memiliki korelasi positif yang sangat tinggi dengan variabel pendidikan, seperti Harapan Lama Sekolah dan Rata-rata Lama Sekolah, yang menandakan bahwa aspek pendidikan menjadi faktor penting dalam meningkatkan kualitas hidup masyarakat.

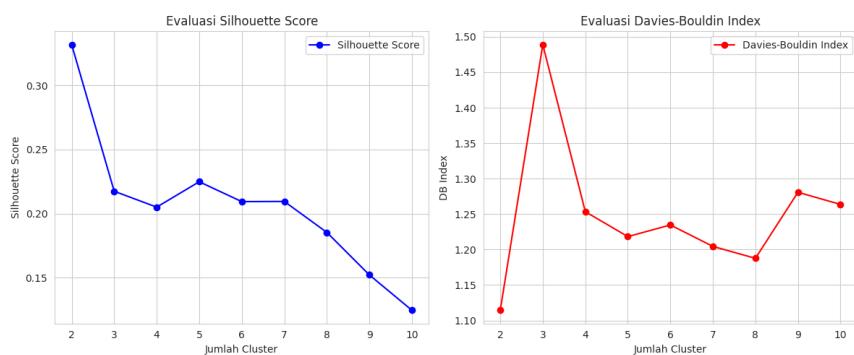
3.3. Preprocessing Data

Pada tahap preprocessing, dilakukan proses transformasi skala terhadap seluruh variabel numerik dengan menggunakan metode RobustScaler dari pustaka scikit-learn. Sebelum proses transformasi, terlebih dahulu kolom non-numerik yaitu 'Kabupaten/Kota' dihapus dari dataframe karena tidak dibutuhkan dalam proses komputasi numerik.

Metode RobustScaler bekerja dengan cara mengurangi nilai median dari setiap fitur dan membaginya dengan interquartile range (IQR), yaitu rentang antara kuartil ke-1 (Q1) dan kuartil ke-3 (Q3). Tujuan dari penggunaan scaling ini adalah untuk menyetarakan skala antar variabel agar tidak terjadi dominasi fitur tertentu dalam proses analisis, khususnya dalam algoritma klasterisasi seperti Spectral Clustering yang sensitif terhadap perbedaan skala antar fitur. Hasil dari proses scaling (X_{scaled}) selanjutnya digunakan sebagai input untuk tahap klasterisasi atau analisis lainnya.

3.4. Evaluasi Jumlah Cluster Optimal

Untuk menentukan jumlah klaster optimal, dilakukan evaluasi dengan Davies-Bouldin Index (DBI) dan Silhouette Score pada skenario jumlah klaster antara 2 hingga 10.



Gambar 3. Grafik Cluster Optimal Berdasarkan Silhouette Score dan Davies-Bouldin Index

Tabel 2. Hasil Evaluasi Cluster Optimal

| Cluster | Nilai Davies Bouldin Index | Nilai Silhouette Score |
|---------|----------------------------|------------------------|
| 2 | 1.1145 | 0.3318 |
| 3 | 1.4889 | 0.2175 |
| 4 | 1.2534 | 0.2051 |
| 5 | 1.2183 | 0.2249 |
| 6 | 1.2347 | 0.2094 |
| 7 | 1.2045 | 0.2096 |
| 8 | 1.1876 | 0.1854 |
| 9 | 1.2807 | 0.1523 |
| 10 | 1.2637 | 0.1247 |

Berdasarkan gambar 3 dan tabel 2 hasil evaluasi menggunakan Silhouette Score dan Davies-Bouldin Index (DBI), jumlah klaster optimal ditentukan pada $k = 2$. Nilai Silhouette Score yang tinggi menunjukkan bahwa objek-objek dalam klaster memiliki kesamaan yang tinggi antar anggotanya dan berbeda dengan klaster lainnya, sedangkan nilai DBI yang rendah menunjukkan bahwa klaster yang terbentuk memiliki kekompakan yang baik serta jarak antar klaster yang cukup jauh. Pada jumlah klaster lebih dari dua, baik nilai Silhouette Score maupun DBI cenderung menunjukkan penurunan kualitas pengelompokan, dengan nilai Silhouette yang semakin rendah dan DBI yang meningkat. Oleh karena itu, pemilihan dua klaster dianggap paling optimal dalam menggambarkan struktur data secara keseluruhan.

3.5. Penerapan Spectral Clustering

- a. Pembangunan Affinity Matrix (Kernel Similarity)

Affinity matrix dibentuk menggunakan Radial Basis Function (RBF) Kernel, yang mengukur kemiripan antar data berdasarkan jarak Euclidean. Parameter gamma diset sebesar 0.1 agar pengaruh outlier menjadi kecil. Nilai gamma yang kecil membuat kernel kurang sensitif terhadap data ekstrem. Matriks hasil (W) ini digunakan sebagai dasar dalam proses Spectral Clustering.

- b. Pembangunan Degree Matrix

Degree matrix merupakan matriks diagonal yang elemen diagonalnya berisi jumlah total kedekatan (similarity) setiap data terhadap data lain.

- c. Perhitungan Normalized Laplacian

Dari affinity matrix dan degree matrix, dihitung normalized Laplacian. Normalized Laplacian adalah bentuk matriks yang digunakan untuk mengekstrak struktur klaster dalam data.

- d. Eigen Decomposition dan Pemilihan Eigenvector

Dari matriks normalized Laplacian (L_{sym}), dilakukan eigen decomposition untuk memperoleh pasangan nilai eigen dan vektor eigen. Selanjutnya, dipilih k vektor eigen terkecil (dalam kasus ini, $k = 2$) sebagai representasi baru dari data. Vektor-vektor ini membentuk ruang fitur baru yang merepresentasikan struktur klaster dalam data, dan akan digunakan pada tahap klasterisasi selanjutnya.

- e. Clustering dengan K-Means

Setelah memperoleh representasi data berdasarkan eigenvector, dilakukan proses klasterisasi menggunakan algoritma K-Means dengan jumlah klaster sebanyak 2.

Hasil dari proses diatas disusun dalam sebuah DataFrame yang berisi nilai eigenvector dan label cluster untuk setiap data, dengan output sebagai berikut:

Tabel 3. Tabel Hasil Reduksi Dimensi dan Label Klaster

| Hasil Eigenvector + Label Cluster: | | | |
|---|---------------|---------------|---------|
| | Eigenvector 1 | Eigenvector 2 | Cluster |
| 0 | -0.738802 | 0.673923 | 1 |
| 1 | -0.978712 | 0.205237 | 1 |
| 2 | -0.914303 | 0.405030 | 1 |

| | | | |
|----|-----------|-----------|---|
| 3 | -0.979439 | -0.201738 | 1 |
| 4 | -0.955551 | 0.294826 | 1 |
| : | : | : | : |
| 37 | -0.654695 | -0.755893 | 0 |

f. Evaluasi Hasil Klaster

Evaluasi klaster dilakukan menggunakan dua metrik: Silhouette Score dan Davies-Bouldin Index. Didapatkan nilai Silhouette Score dan nilai Davies-Bouldin Index terbaik berdasarkan uji coba gamma sebagai berikut.

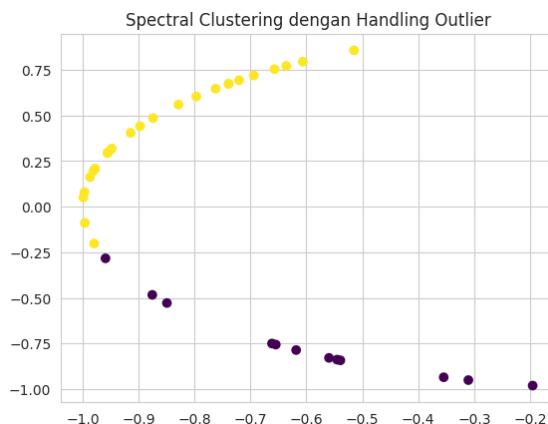
Tabel 4. Hasil Evaluasi Kinerja Model Klasterisasi Menggunakan Metrik Silhouette Score dan Davies-Bouldin Index

| Keterangan Metriks Evaluasi | Nilai Evaluasi |
|-----------------------------|----------------|
| Silhouette Score | 0.6655 |
| Davies-Bouldin Index | 0.4401 |

Hasil evaluasi clustering pada tabel 4 menunjukkan bahwa kualitas pengelompokan yang dihasilkan tergolong baik. Nilai Silhouette Score sebesar 0.6655 termasuk dalam kategori tinggi, yang mengindikasikan bahwa data dalam tiap cluster cukup seragam dan memiliki pemisahan yang jelas antar kelompok. Skor ini mencerminkan struktur klaster yang solid dan terdefinisi dengan baik. Sementara itu, nilai Davies-Bouldin Index (DBI) sebesar 0.4401 juga tergolong baik, karena berada di bawah ambang batas 0.5. Hal ini menunjukkan bahwa masing-masing cluster memiliki kekompakan internal yang tinggi dan terpisah dengan baik dari cluster lainnya.

Secara keseluruhan, kombinasi Silhouette Score yang tinggi dan DBI yang rendah menunjukkan bahwa hasil clustering memiliki kualitas yang kuat dan pengelompokan yang terbentuk cukup optimal.

g. Visualisasi Hasil Klasterisasi



Gambar 4. Visualisasi Hasil Klasterisasi

Visualisasi hasil Spectral Clustering pada gambar 4 dengan penanganan outlier menunjukkan bahwa data berhasil dikelompokkan menjadi dua klaster yang terpisah dengan jelas. Sebagian besar data tergabung dalam satu klaster utama (berwarna kuning) yang membentuk pola melengkung dan rapat, menunjukkan kemiripan karakteristik antar anggota klaster. Sementara itu, sebagian kecil data berada di klaster lain (berwarna ungu) yang terpisah jauh, mengindikasikan adanya data yang berbeda secara signifikan, kemungkinan merupakan outlier. Pemisahan ini menunjukkan bahwa metode *handling outlier* yang diterapkan berhasil meningkatkan kualitas pengelompokan dengan membedakan data utama dari data yang menyimpang.

3.6. Pemetaan Hasil Cluster

Setelah dilakukan klasterisasi menggunakan Spectral Clustering, diperoleh pembagian wilayah menjadi 2 klaster berdasarkan variabel indikator kemiskinan. Tabel 5 menyajikan daftar kabupaten/kota di Provinsi Jawa Timur beserta hasil klasterisasi yang diperoleh berdasarkan penerapan metode yang digunakan.

Tabel 5. Hasil Klasterisasi Kabupaten/Kota di Jawa Timur Berdasarkan Indikator Kemiskinan

| No | Kabupaten/Kota | Cluster |
|----|------------------|---------|
| 1 | Sidoarjo | 0 |
| 2 | Kota Malang | 0 |
| 3 | Kota Blitar | 0 |
| 4 | Kota Kediri | 0 |
| 5 | Gresik | 0 |
| 6 | Lamongan | 0 |
| 7 | Kota Pasuruan | 0 |
| 8 | Kota Probolinggo | 0 |
| 9 | Kota Madiun | 0 |
| 10 | Kota Mojokerto | 0 |
| 11 | Kota Batu | 0 |
| 12 | Kota Surabaya | 0 |
| 13 | Lumajang | 1 |
| 14 | Jember | 1 |
| 15 | Trenggalek | 1 |
| 16 | Pacitan | 1 |
| 17 | Mojokerto | 1 |
| 18 | Pasuruan | 1 |
| 19 | Probolinggo | 1 |
| 20 | Situbondo | 1 |
| 21 | Banyuwangi | 1 |
| 22 | Jombang | 1 |
| 23 | Madiun | 1 |
| 24 | Nganjuk | 1 |
| 25 | Tuban | 1 |
| 26 | Bojonegoro | 1 |
| 27 | Bondowoso | 1 |
| 28 | Ponorogo | 1 |
| 29 | Blitar | 1 |
| 30 | Kediri | 1 |
| 31 | Malang | 1 |
| 32 | Tulungagung | 1 |
| 33 | Sumenep | 1 |
| 34 | Bangkalan | 1 |
| 35 | Sampang | 1 |
| 36 | Pamekasan | 1 |
| 37 | Magetan | 1 |
| 38 | Ngawi | 1 |

Tabel 5 menunjukkan hasil pengelompokan Kabupaten/Kota di Jawa Timur berdasarkan tingkat kemiskinan. Klaster 0 terdiri dari 12 daerah dengan tingkat kemiskinan rendah, yang sebagian besar

adalah kota-kota besar seperti Kota Surabaya, Kota Malang, dan Sidoarjo. Klaster ini menandakan daerah-daerah yang cenderung lebih maju secara ekonomi dan memiliki kondisi sosial yang lebih baik.

Sedangkan klaster 1 berisi 26 Kabupaten/Kota dengan tingkat kemiskinan tinggi. Daerah-daerah dalam klaster ini, seperti Banyuwangi, Jember, dan Sumenep, umumnya menghadapi tantangan sosial ekonomi yang lebih besar dan memerlukan perhatian khusus dalam upaya pengentasan kemiskinan.

3.7. Analisis Hasil

a. Rata – Rata Variabel Berdasarkan Klaster

Klaster 0 terdiri dari daerah dengan tingkat kemiskinan rendah, rata-rata persentase penduduk miskin hanya 6,68% dan jumlah penduduk miskin sekitar 56,44 ribu jiwa. Daerah-daerah ini juga memiliki angka harapan hidup yang tinggi (73,63 tahun), rata-rata lama sekolah lebih panjang (10,36 tahun), serta Indeks Pembangunan Manusia (IPM) tinggi sebesar 79,63. Upah pekerja informal juga tinggi, rata-rata Rp1,92 juta/bulan, yang mencerminkan aktivitas ekonomi yang lebih baik. Sebaliknya, Klaster 1 dihuni oleh daerah dengan tingkat kemiskinan tinggi, yaitu rata-rata persentase penduduk miskin 12,01% dan jumlah penduduk miskin mencapai 135,06 ribu jiwa. Daerah-daerah ini memiliki akses pendidikan lebih rendah (rata-rata lama sekolah hanya 7,46 tahun), angka harapan hidup lebih rendah (71,86 tahun), dan IPM yang lebih rendah (70,94). Rata-rata penghasilan pekerja informal juga lebih kecil, hanya Rp1,51 juta/bulan.

Meskipun analisis utama menggunakan nilai rata-rata, pendekatan ini juga mempertimbangkan sebaran data antarwilayah dengan memperhatikan pola variasi nilai minimum dan maksimum dari masing-masing variabel. Hal ini penting karena persentase penduduk miskin dapat dipengaruhi oleh jumlah penduduk absolut, sehingga interpretasi hasil tidak hanya didasarkan pada nilai rata-rata semata, melainkan juga mempertimbangkan konteks ukuran wilayah dan karakteristik sosial ekonomi yang berbeda. Kategorisasi wilayah dalam tingkat kemiskinan tinggi dan rendah ditentukan berdasarkan kombinasi beberapa indikator sosial ekonomi utama seperti garis kemiskinan, rata-rata lama sekolah, dan IPM. Pendekatan ini sejalan dengan konsep kemiskinan multidimensi yang menekankan keterkaitan antara faktor pendidikan, kesehatan, dan daya beli masyarakat.

b. Feature Importance

Berdasarkan hasil analisis feature importance terhadap model clustering kemiskinan di kabupaten/kota Jawa Timur, diperoleh informasi mengenai variabel-variabel yang paling berkontribusi dalam membentuk pengelompokan wilayah miskin dan tidak miskin.

Tabel 6. Bobot Penting Setiap Fitur dalam Proses Klasterisasi Menggunakan Spectral Clustering

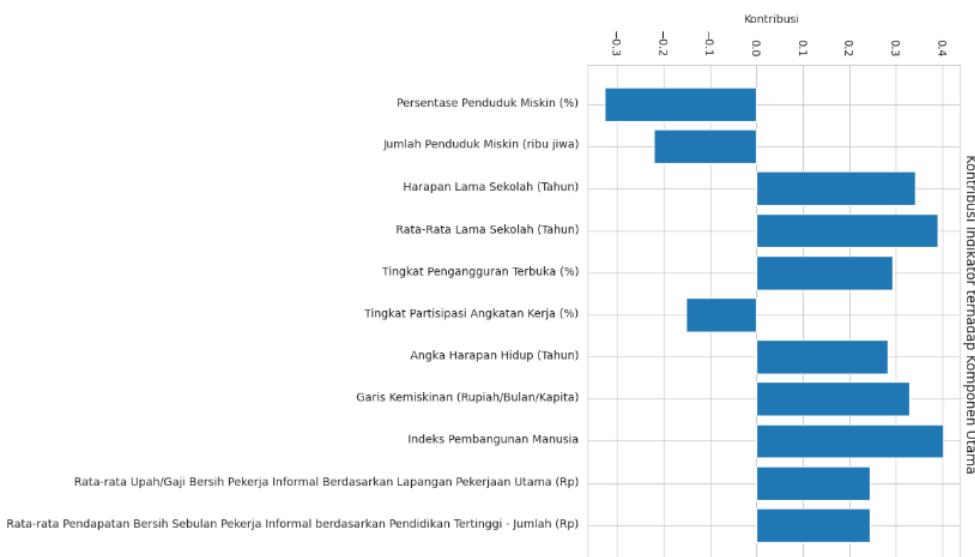
| No | Fitur | Importance |
|----|---|------------|
| 1 | Garis Kemiskinan (Rupiah/Bulan/Kapita) | 0.259418 |
| 2 | Rata-Rata Lama Sekolah (Tahun) | 0.259067 |
| 3 | Harapan Lama Sekolah (Tahun) | 0.147613 |
| 4 | Indeks Pembangunan Manusia | 0.133441 |
| 5 | Persentase Penduduk Miskin (%) | 0.061056 |
| 6 | Jumlah Penduduk Miskin (ribu jiwa) | 0.044469 |
| 7 | Rata-rata Upah/Gaji Bersih Pekerja Informal (Lapangan Utama) (Rp) | 0.033852 |
| 8 | Rata-rata Pendapatan Bersih Sebulan Pekerja Informal (Pendidikan) | 0.030777 |
| 9 | Tingkat Pengangguran Terbuka (%) | 0.013731 |
| 10 | Angka Harapan Hidup (Tahun) | 0.009560 |
| 11 | Tingkat Partisipasi Angkatan Kerja (%) | 0.007017 |

Berdasarkan analisis feature importance dengan algoritma Random Forest pada Tabel 6, variabel paling berpengaruh dalam pembentukan klaster kemiskinan di Jawa Timur adalah Garis Kemiskinan

(0.259418) dan Rata-Rata Lama Sekolah (0.259067), yang menunjukkan peran dominan aspek ekonomi dan pendidikan. Disusul oleh Harapan Lama Sekolah (0.147613) dan Indeks Pembangunan Manusia (0.133441), yang juga berpengaruh signifikan. Sementara itu, variabel seperti Persentase dan Jumlah Penduduk Miskin berpengaruh lebih kecil, dan indikator seperti Tingkat Pengangguran Terbuka, Angka Harapan Hidup, serta Partisipasi Angkatan Kerja menunjukkan kontribusi yang rendah dalam pembentukan klaster.

Hasil feature importance pada penelitian ini merepresentasikan tingkat kontribusi masing-masing variabel terhadap proses pembentukan klaster, bukan menunjukkan hubungan sebab-akibat antarvariabel. Oleh karena itu, hasil analisis ini bersifat deskriptif dan digunakan untuk mengidentifikasi indikator-indikator yang berperan dominan dalam struktur pengelompokan wilayah, tanpa menguji signifikansi pengaruh secara statistik.

- Visualisasi kontribusi terhadap komponen utama dengan PCA

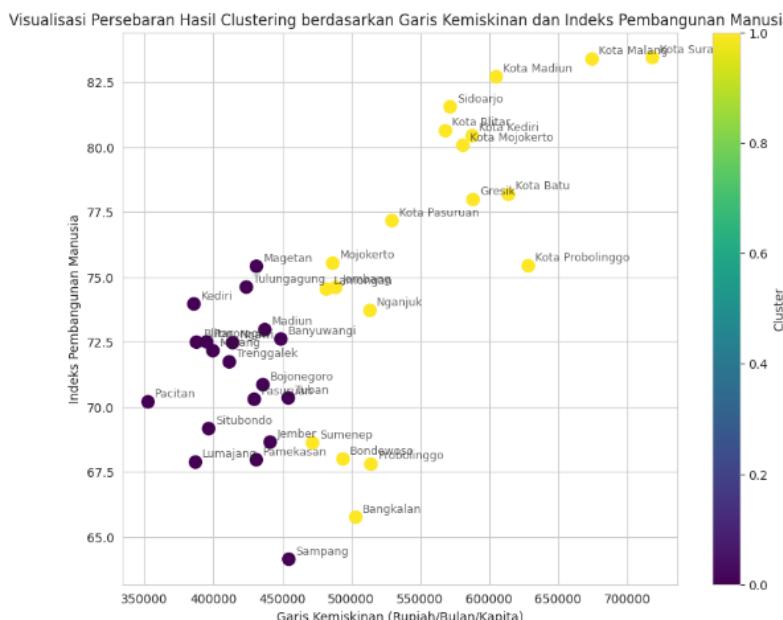


Gambar 5. Analisis PCA dari Indikator yang Berpengaruh

Gambar 5 menunjukkan kontribusi masing-masing indikator terhadap komponen utama dalam PCA. Indikator seperti Indeks Pembangunan Manusia, Harapan Lama Sekolah, Rata-Rata Lama Sekolah, dan Garis Kemiskinan memberikan kontribusi positif besar, yang menunjukkan bahwa wilayah dengan nilai tinggi pada indikator ini cenderung memiliki tingkat kemiskinan lebih rendah. Sebaliknya, Persentase dan Jumlah Penduduk Miskin berkontribusi negatif, mengarah pada klasifikasi wilayah miskin. Indikator lain seperti Angka Harapan Hidup, Partisipasi Angkatan Kerja, dan Tingkat Pengangguran memiliki kontribusi lebih kecil. Secara keseluruhan, aspek pendidikan dan daya beli menjadi faktor utama dalam membedakan kelompok kemiskinan wilayah.

- Visualisasi Spectral Clustering Berdasarkan 2 Variabel Utama

Berdasarkan visualisasi hasil Spectral Clustering yang memanfaatkan dua variabel utama yaitu Garis Kemiskinan (Rupiah/Bulan/Kapita) dan Indeks Pembangunan Manusia (IPM), kabupaten/kota di Jawa Timur berhasil dikelompokkan ke dalam dua cluster yang berbeda.



Gambar 6. Visualisasi Spectral Clustering Berdasarkan 2 Variabel Utama

Gambar 6 memperlihatkan hasil clustering Spectral Clustering berdasarkan variabel Garis Kemiskinan dan IPM, yang menghasilkan dua cluster. Cluster 0 (ungu) mencakup wilayah dengan garis kemiskinan rendah namun IPM masih rendah hingga sedang, seperti Sampang dan Pacitan. Sementara itu, cluster 1 (kuning) berisi daerah dengan IPM tinggi dan garis kemiskinan yang juga tinggi, seperti Kota Surabaya dan Sidoarjo, mencerminkan wilayah dengan kualitas hidup dan pembangunan yang lebih baik.

e. Implikasi Kebijakan

Hasil penelitian ini dapat menjadi dasar dalam merumuskan kebijakan intervensi yang lebih terfokus pada wilayah dengan tingkat kerentanan tinggi. Misalnya, daerah yang termasuk dalam klaster kemiskinan tinggi dapat menjadi prioritas dalam program peningkatan akses pendidikan melalui bantuan sarana sekolah, pelatihan keterampilan kerja, serta program peningkatan kualitas tenaga kerja informal. Selain itu, intervensi dapat difokuskan pada peningkatan daya beli masyarakat melalui penguatan ekonomi lokal, seperti pemberdayaan UMKM dan perluasan akses permodalan. Dengan demikian, hasil klasterisasi tidak hanya memberikan gambaran tingkat kemiskinan antarwilayah, tetapi juga dapat menjadi acuan dalam menentukan arah kebijakan pembangunan yang lebih tepat sasaran.

4. KESIMPULAN

Penelitian ini bertujuan untuk memetakan tingkat kemiskinan di kabupaten/kota di Jawa Timur dengan menggunakan metode Spectral Clustering berdasarkan 11 indikator sosial ekonomi. Hasil penelitian menunjukkan bahwa wilayah dapat dikelompokkan ke dalam dua klaster utama, yaitu klaster dengan tingkat kemiskinan tinggi dan klaster dengan tingkat kemiskinan rendah. Wilayah dengan tingkat kemiskinan tinggi ditandai oleh rendahnya angka harapan hidup, pengeluaran per kapita, serta tingginya persentase penduduk miskin. Sebaliknya, wilayah dengan tingkat kemiskinan rendah memiliki indikator sosial ekonomi yang lebih baik, terutama dalam aspek pendidikan, kesehatan, dan kesejahteraan.

Metode Spectral Clustering terbukti efektif dalam mengelompokkan data multidimensional tanpa memerlukan inisialisasi centroid, dengan nilai Davies-Bouldin Index sebesar 0,4401 dan Silhouette

Score sebesar 0,6655. Hasil ini memberikan dasar bagi pemerintah daerah untuk merancang kebijakan pengentasan kemiskinan yang lebih fokus berdasarkan karakteristik masing-masing klaster.

Sebagai tindak lanjut, penelitian selanjutnya disarankan untuk memperluas cakupan analisis dengan memasukkan variabel tambahan seperti karakteristik geografis, akses terhadap infrastruktur dasar, serta ketersediaan sarana dan prasarana publik guna memperoleh gambaran kemiskinan yang lebih komprehensif. Selain itu, penggunaan data panel atau time-series dapat dilakukan untuk menganalisis dinamika perubahan klaster dari waktu ke waktu. Pendekatan lain seperti integrasi metode optimasi (misalnya Particle Swarm Optimization) atau model prediktif berbasis machine learning juga dapat diterapkan untuk meningkatkan akurasi pengelompokan serta mendukung perumusan kebijakan intervensi yang lebih tepat sasaran.

DAFTAR PUSTAKA

- [1] N. Shalsadilla, S. Martha, H. Perdana, and E. Sulistianingsih, “Penentuan Jumlah Cluster Optimum Menggunakan Davies Bouldin Index dalam Pengelompokan Wilayah Kemiskinan di Indonesia,” vol. 23, no. 1, pp. 63–72, 2023.
- [2] D. Suhartono, S. Kanto, and S. Mu, “Kajian Tentang Makna , Penyebab , dan Strategi Penanggulangan Kemiskinan (Kajian Fenomenologi Komunitas Pemulung di TPA Supit Urang Kota Malang),” *Wacana*, vol. 17, no. 4, pp. 234–245, 2014.
- [3] T. M. Fahrudin *et al.*, “Cluster Analysis of Hospital Inpatient Service Efficiency Based on BOR, BTO, TOI, AvLOS Indicators using Agglomerative Hierarchical Clustering,” vol. 18, no. 2, pp. 194–210, 2021, doi: 10.31515/telematika.v18i2.4786.
- [4] M. W. Putri, I. M. Nur, and R. Wasono, “Implementasi Spectral Clustering Algorithm Untuk Pengelompokan Sasaran Vaksinasi Covid-19 Di Indonesia,” *J. Stat. Univ. Muhammadiyah Semarang*, vol. 10, no. 1, p. 26, 2022, doi: 10.26714/jsunimus.10.1.2022.26-31.
- [5] F. Alfiah, A. Almadayani, D. Al Farizi, and E. Widodo, “Analisis Clustering K-Medoids Berdasarkan Indikator Kemiskinan di Jawa Timur Tahun 2020,” *J. Ilm. Sains*, vol. 22, no. 1, p. 1, 2021, doi: 10.35799/jis.v22i1.35911.
- [6] A. Yusuf and H. Tjandrasa, “Prediksi Nilai dengan Metode Spectral Clustering dan Clusterwise Regression,” *J. ScimateC*, vol. 4, no. 1, pp. 1–8, 2014.
- [7] A. Rahmasari, “Pengaruh Tingkat Pendidikan , Kepadatan Penduduk , dan Tingkat Pengangguran Terbuka Terhadap Tingkat Kemiskinan di Kabupaten Jombang Tahun 2011-2015,” vol. 3, pp. 73–81, 2023, doi: 10.46821/bharanomics.v3i2.574.
- [8] M. Radhi, A. Amalia, D. R. H. Sitompul, S. H. Sinurat, and E. Indra, “Analisis Big Data Dengan Metode Exploratory Data Analysis (Eda) Dan Metode Visualisasi Menggunakan Jupyter Notebook,” *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 4, no. 2, pp. 23–27, 2022, doi: 10.34012/journalsisteminformasidanilmukomputer.v4i2.2475.
- [9] S. S. M. Wara, A. F. Adziima, M. Nasrudin, and A. R. Pratama, “Predictive Analysis of Government Application Comment on Playstore with Clustered Support Vector Machine,” *Proceeding - IEEE 10th Inf. Technol. Int. Semin. ITIS 2024*, pp. 84–88, 2024, doi: 10.1109/ITIS64716.2024.10845453.
- [10] I. Zahy, A. Illah, W. Syaifullah, J. Sapu, and A. T. Damaliana, “Implementasi Metode Klasifikasi LightGBM dan Analisis Survival dalam Memprediksi Pelanggan Churn,” vol. 8, no. 1, pp. 43–53, 2024.
- [11] S. Renaldi, S. D. A. Prasetya, and A. Muhammin, “Analisis Klaster Partitioning Around Medoids dengan Gower Distance untuk Rekomendasi Indekos (Studi Kasus: Indekos di Sekitar Kampus UPNVJT),” *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 2060–2069, 2024, doi: 10.33379/gtech.v8i3.4898.
- [12] A. Matz, “Customer Loyalty Clustering Model Using K-Means Algorithm with LRIFMQ Parameters,” *J. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 5, no. 2, p. 54, 2020.
- [13] D. A. Prasetya, A. P. Sari, M. Idhom, and A. Lisanthoni, “Optimizing Clustering Analysis to Identify High-Potential Markets for Indonesian Tuber Exports,” vol. 7, no. 1, pp. 113–122, 2025.
- [14] S. Wulandari and D. Novita, “Analisis Clustering Virus MERS-CoV Menggunakan Metode Spectral Clustering Dan Algoritma K-Means,” *STRING (Satuan Tulisan Ris. dan Inov. Teknol.)*, vol. 5, no. 3, p. 315, 2021, doi: 10.30998/string.v5i3.7942.

- [15] M. Sukarno Hatta, F. Azmi, and C. Setianingsih, “Clustering Pada Data Sentimen Penggunaan Transportasi Online Menggunakan Algoritma Spectral Clustering,” *e-Proceeding Eng.*, vol. 8, no. 6, p. 11945, 2021.
- [16] A. R. Pratama, B. S. B. Dewantara, D. M. Sari, and D. Pramadihanto, “Improvement of DBSCAN Algorithm Involving Automatic Parameters Estimation and Curvature Analysis in 3D Point Cloud of Piled Pipe,” *J. Image Graph. Kingdom*, vol. 12, no. 2, pp. 175–185, 2024, doi: 10.18178/joig.12.2.175-185.