Jurnal Ilmu Komputer Revolusioner

Vol.8 No.10, Oktober 2024
ISSN: 24430355

ANALISIS HARGA RUMAH DI DAERAH JAKARTA SELATAN DENGAN MENGGUNAKAN ALGORITMA SPECTRAL CLUSTERING

Reo Rizki Ananda[1], Arnita[2], Raysa Maharani[3], Mufida Azza Iskandar Lubis[4], Sigun Putra Hasian Sitompul[5]
Universitas Negeri Medan
E-mail: reorizki07@gmail.com[1], arnita@gmail.com[2], raysa28maharani@gmail.com[3], mufidaazza80@gmail.com[4], sigunsitompul1@gmail.com[5]

Abstract

The analysis of house prices in South Jakarta was conducted using the Spectral Clustering method, which is known for its ability to handle complex and non-linear data. Factors such as location, amenities, land area, and economic conditions influence property prices, creating patterns that are difficult to identify with traditional clustering methods. Spectral Clustering utilizes graph theory and spectral decomposition to more accurately group property data. In this study, a dataset was used that includes variables such as house prices, land area, building area, number of bedrooms, bathrooms, and the availability of a garage. The process began with data cleaning, followed by the application of the Spectral Clustering algorithm. The clustering results were evaluated using the Silhouette Score and Davies-Bouldin Index. The analysis results indicate that properties in South Jakarta can be optimally grouped into two to three clusters. Most houses fall into a cluster with moderate prices and sizes, while properties with high prices and large land sizes form a separate cluster.

Keywords *— Preprocessing, House Prices, South Jakarta, Spectral Clustering, Property Analysis*
.

Abstrak

Analisis harga rumah di Jakarta Selatan dilakukan dengan menggunakan metode Spectral Clustering, yang dikenal mampu menangani data yang kompleks dan non-linear. Faktor-faktor seperti lokasi, fasilitas, luas tanah, dan kondisi ekonomi memengaruhi harga properti, menciptakan pola yang sulit diidentifikasi oleh metode clustering tradisional. Spectral Clustering memanfaatkan teori graf dan dekomposisi spektral untuk mengelompokkan data properti secara lebih tepat. Dalam studi ini, digunakan dataset yang mencakup variabel harga rumah, luas tanah, luas bangunan, jumlah kamar tidur, kamar mandi, serta ketersediaan garasi. Proses dimulai dengan pembersihan data, diikuti penerapan algoritma Spectral Clustering. Evaluasi hasil klasterisasi dilakukan menggunakan Silhouette Score dan Davies-Bouldin Index. Hasil analisis menunjukkan bahwa properti di Jakarta Selatan dapat dikelompokkan secara optimal ke dalam dua hingga tiga klaster. Sebagian besar rumah berada dalam kelompok harga dan ukuran moderat, sementara properti dengan harga tinggi dan ukuran lahan besar terbentuk sebagai klaster terpisah.

Kata Kunci — Preprocessing, Harga Rumah, Jakarta Selatan, Spectral Clustering, Analisis Properti.

1. PENDAHULUAN

Rumah adalah kebutuhan pokok bagi setiap individu karena berperan sebagai tempat tinggal yang memberikan rasa aman dan nyaman. Selain itu, rumah juga menjadi tempat berkumpulnya keluarga serta menjalankan aktivitas sehari-hari. Dari segi harga, memiliki rumah yang terjangkau sangatlah penting karena mempengaruhi kemampuan orang untuk

memiliki hunian sendiri [1]. Fasilitas yang ada di dalam rumah, seperti air bersih, listrik, sanitasi yang baik, serta akses ke transportasi dan layanan publik, juga sangat berpengaruh. Rumah dengan harga yang sesuai dan fasilitas lengkap mampu meningkatkan kualitas hidup penghuninya secara signifikan [2] [3].

Sektor properti di Indonesia telah menunjukkan peningkatan signifikan dalam beberapa tahun terakhir. Salah satu aspek penting dalam sektor ini adalah harga rumah, yang menjadi pertimbangan utama bagi calon pembeli. Di Daerah Jakarta Selatan, pembangunan perumahan tumbuh dengan cepat. Faktor-faktor seperti pertumbuhan ekonomi, urbanisasi, dan perkembangan infrastruktur di wilayah Jakarta Selatan telah menyebabkan peningkatan permintaan rumah. Oleh karena itu, sangat penting memiliki alat pengelompokan yang andal untuk memprediksi data harga rumah di wilayah tersebut [4].

Analisis harga properti, khususnya rumah, merupakan aspek yang sangat penting dalam berbagai bidang seperti ekonomi, perencanaan kota, dan manajemen properti. Harga rumah di suatu daerah sering kali dipengaruhi oleh berbagai faktor yang saling berkaitan, termasuk lokasi rumah, kemudahan akses ke berbagai fasilitas, ketersediaan fasilitas umum, serta kondisi ekonomi di sekitar wilayah tersebut. Dalam penelitian ini, akan melakukan analisis terhadap data penjualan rumah yang terletak di Jakarta Selatan. Metode yang digunakan dalam analisis ini adalah spectral clustering, sebuah teknik yang dapat membantu mengidentifikasi pola atau kelompok dalam data yang memiliki kesamaan tertentu.

Kasus analisis harga rumah dengan metode spectral Clustering perlu diangkat dalam penelitian karena metode ini menawarkan pendekatan yang akurat dalam melakukan pengelompokkan data yang kompleks. Dapat diketahui konteks harga rumah, berbsgai faktor seperti lokasi, fasilitas maupun pertumbuhan ekonomi akan menciptakan pola yang tidak dapat di identifikasi. Selain itu, kasus ini diangkat agar dapat memberikan kontribusi signifikan dalam memahami dinamika pasar properti.

Dalam beberapa dekade terakhir, spectral clustering telah berkembang menjadi solusi yang efektif untuk menangani data yang kompleks. Metode ini unggul dalam mengidentifikasi pola non-konveks, yang sangat berguna dalam analisis harga rumah, di mana faktor geografis dan sosio-ekonomi sering membentuk kluster yang tidak teratur. Spectral clustering memadukan prinsip-prinsip teori graf dan dekomposisi spectral untuk memproyeksikan data ke ruang dimensi lebih rendah, sehingga mempermudah proses pengelompokan [5] [6].

Metode ini juga bekerja dengan membangun sebuah grafik dari data yang ada. Dalam grafik ini, setiap titik (vertex) mewakili satu record atau data rumah, sedangkan garis penghubung (edge) antara titik-titik tersebut menunjukkan keterhubungan atau hubungan antara data rumah. Keterhubungan ini biasanya diukur berdasarkan jarak atau kemiripan antara data rumah [7].

Untuk mengevaluasi hasil clustering, digunakan tiga metrik evaluasi: Silhouette Score, Davies-Bouldin Index, dan Calinski-Harabasz Score. Silhouette Score mengukur seberapa mirip objek dalam satu klaster dibandingkan dengan objek di klaster lain, dengan nilai berkisar antara -1 hingga 1, di mana nilai yang lebih tinggi menunjukkan klasterisasi yang lebih baik. Davies-Bouldin Index mengevaluasi kompaksi dan pemisahan antar klaster, dengan nilai yang lebih rendah menunjukkan klaster yang lebih baik. Calinski-Harabasz Score menilai rasio antara jumlah total dispersi antar klaster dengan jumlah total dispersi intra-klaster, di mana nilai yang lebih tinggi menunjukkan klaster yang lebih baik.

Silhouette Score memberikan ukuran seberapa baik setiap titik data ditempatkan

dalam Cluster mereka sendiri dibandingkan dengan Cluster lain, sementara Davies-Bouldin Index mengevaluasi seberapa jauh Cluster tersebut berada dari yang lain.Kedua metrik ini memberikan informasi yang berharga tentang kualitas dan kehomogenan Cluster yang dihasilkan oleh algoritma Clustering [8].

Indeks Davies-Bouldin (DBI) adalah salah satu metode yang digunakan untuk mengukur validitas kluster dalam suatu metode pengelompokan. Pengukuran dengan Indeks Davies-Bouldin ini memaksimalkan jarak antar-kluster dan pada saat yang sama berusaha meminimalkan jarak antara titik-titik dalam satu kluster. Jika jarak antar-kluster maksimal, itu berarti karakteristik setiap kluster sedikit mirip, sehingga perbedaan antar-kluster terlihat lebih jelas. Jika jarak intra-kluster minimal, itu berarti setiap objek dalam kluster memiliki tingkat kesamaan karakteristik yang tinggi. Hasil pengelompokan yang diperoleh dari penentuan titik pusat kluster yang diusulkan kemudian dievaluasi menggunakan metode DBI. Sehingga dapat dilihat korelasi metode penentuan titik pusat kluster berdasarkan Sum of Squared Error terhadap peningkatan kualitas kluster berdasarkan nilai DBI yang diperoleh [9].

Pada tahun 1974, Calinski T dan Harabasz J mengusulkan Indeks Calinski-Harabasz (CH), yang mengevaluasi efek klustering berdasarkan kekompakan kluster dan kekompakan antar kluster [10] dan dikenal sebagai indeks Variasi Rasio, adalah salah satu ukuran untuk mengevaluasi kualitas klustering. Indeks ini mengukur seberapa baik kluster yang terbentuk terpisah dari satu sama lain dan seberapa kompak elemen di dalam kluster tersebut.

Banyak penelitian sebelumnya telah menggunakan berbagai metode clustering dalam menganailisis harga rumah. Salah satu metode paling populer sebelum Spectral Clustering ialah K-means Clustering. Namun K-Means Clustering memiliki batasan dalam mengatasi kluster yang berbentuk tidak koveks dan data heterogen. [11] [12] Akan tetapi, Spectral Clustering salah satu metode clustering yang mampu dalam menangani kluster non-konveks dan data yang tidak terdistribusi secara linier [13].

Spectral Clustering juga algoritma pengelompokan yang berkembang yang telah berkinerja lebih baik daripada banyak algoritma pengelompokan tradisional dalam banyak kasus. Spectral Clustering membuat setiap titik data sebagai grafik-node dan kemudian mengubah masalah pengelompokan menjadi masalah partisi-grafik. Selain itu keunggulan Spectral Clustering, Ssalah teknik pengelompokan yang sangat user-friendly dan relatif cepat, terutama untuk data yang memiliki dimensi tinggi, bahkan hingga data berjumlah ribuan. Dalam metode spectral clustering, pendekatan ini menganggap pengelompokan data sebagai grafik dan melakukan partisi data tanpa asumsi tertentu mengenai bentuk klaster data [14] [15].

Maka dari itu, metode Spectral Clustering pada penelitian ini sangat penting, dikarenakan Spectral Clustering memiliki kemampuan dalam mengidentifikasi pola non-liniear dan kluster yang kompleks berdasarkan faktor yang mempengaruhi pasar properti.

2. METODE PENELITIAN

Penelitian ini menggunakan data harga rumah di Jakarta Selatan yang terdiri dari beberapa variabel: harga rumah (Rp), luas tanah (m^2), luas bangunan (m^2), jumlah kamar tidur, jumlah kamar mandi, dan ketersediaan garasi. Dataset yang digunakan diambil dari beberapa situs penjualan properti, seperti rumah123.com, dan mencakup 1000 properti pada tahun 2020 :

Tabel 1 Data Harga Rumah

Harga(Rp)	Luas Tanah (m^2)	Luas Bangunan (m^2)	Jumlah kamar Tidur	Jumlah Kamar Mandi	Garasi

Count	920000000	920	920	920	920	920
Mean	12979850000000000	448.74	424.42	4.33	3.81	0.76
Std	8123437000000000	336.43	216.68	1.73	1.68	0.42
Min	430000000000000	22	38	1	1	0
25%	6500000000000000	210	285	4	3	1
50%	11825000000000000	380	400	4	4	1
75%	18000000000000000	585.5	530	5	4	1
Max	36000000000000000	3000	1944	27	27	1

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan algoritma Spectral Clustering untuk mengelompokkan harga rumah di Jakarta Selatan berdasarkan beberapa variabel, seperti harga rumah, luas tanah, dan luas bangunan. Dari analisis yang dilakukan, ditemukan bahwa metode ini mampu mengidentifikasi dua klaster utama yang paling optimal dalam dataset property.

1. Hasil Klasterisasi

Dari hasil implementasi Spectral Clustering, diperoleh bahwa pembagian dua klaster memberikan hasil klasterisasi yang paling optimal, sebagaimana ditunjukkan oleh Silhouette Score tertinggi dengan nilai 0,82. Berikut adalah dua klaster utama yang teridentifikasi:

1. Klaster 1: Berisi mayoritas properti dengan harga di bawah 2 x 10^{10} IDR dan luas tanah kurang dari 1000 m^2. Properti dalam klaster ini memiliki karakteristik yang cenderung lebih moderat.
2. Klaster 2: Klaster ini berisi properti yang memiliki harga dan luas tanah yang jauh lebih besar dari klaster pertama. Properti-properti ini teridentifikasi sebagai outlier, yang menunjukkan properti mewah dengan karakteristik unik.

Penambahan jumlah klaster menjadi tiga atau lebih menyebabkan penurunan signifikan pada Silhouette Score, yang turun di bawah 0,4. Hal ini menunjukkan bahwa dua klaster sudah cukup optimal untuk menggambarkan karakteristik pasar properti di Jakarta Selatan.

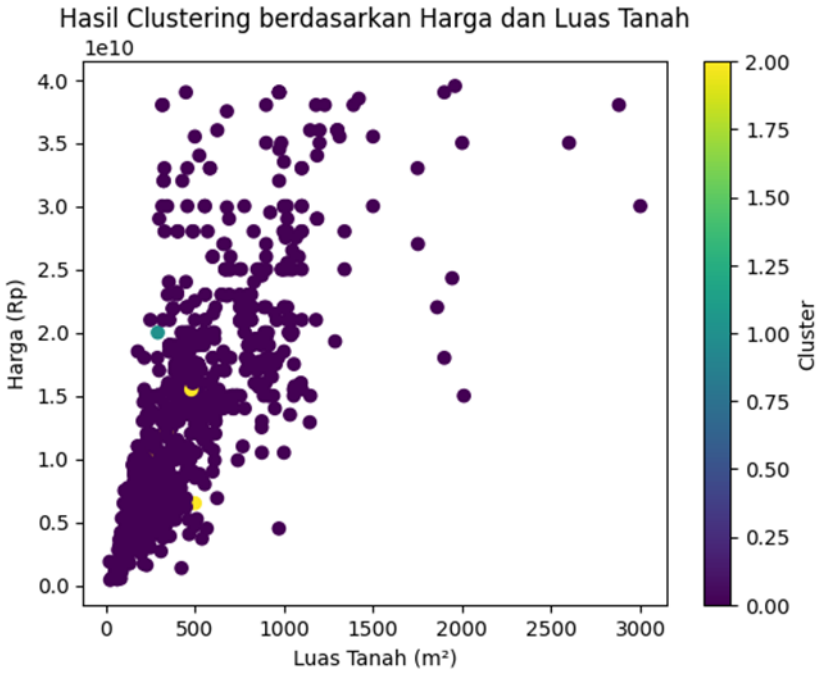
Gambar 1 Silhoutte Scoe vs Number of Clusters

2. Visualisasi Hasil Klasterisasi

Scatter plot yang dihasilkan dari analisis ini memperlihatkan distribusi harga rumah. Boxplot menunjukkan bahwa sebagian besar rumah dijual dengan harga sekitar 1,25 x 10^{10} IDR, dengan beberapa outlier yang mencapai harga hingga 4,0 x 10^{10} IDR. Sedangkan, scatter plot yang memetakan luas tanah terhadap harga rumah menunjukkan adanya perbedaan mencolok antara rumah dengan harga tinggi dan rendah, dengan klaster yang terpisah jelas.

Tabel 2 Groupping cluster dengan rata-rata

Cluster	Jumlah Cluster	Harga(Rp)	Luas Tanah (m^2)	Luas Bangunan (m^2)	Jumlah kamar Tidur	Jumlah Kamar Mandi	Garasi
0	930	13424850000000000	461.63	431.98	4.25	3.74	0.76
1	4	13675000000000000	246	622	27	27	0.5
2	2	11125000000000000	387.5	475	15	12.25	1



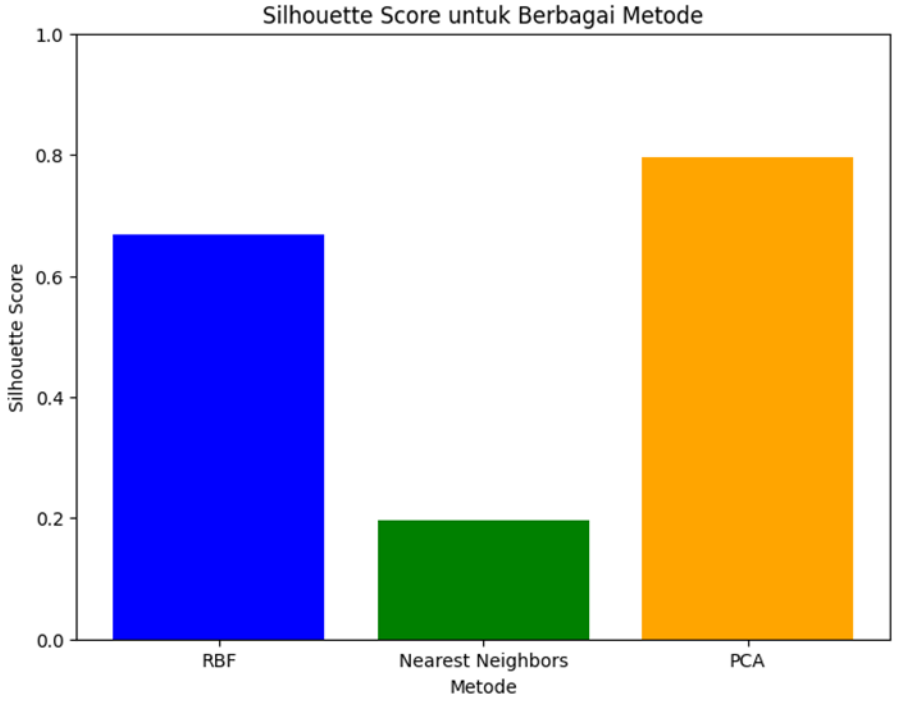
Gambar 2 Hasil Clustering Berdasarkan Harga dan Luas Tanah

3. Evaluasi Kualitas Klasterisasi

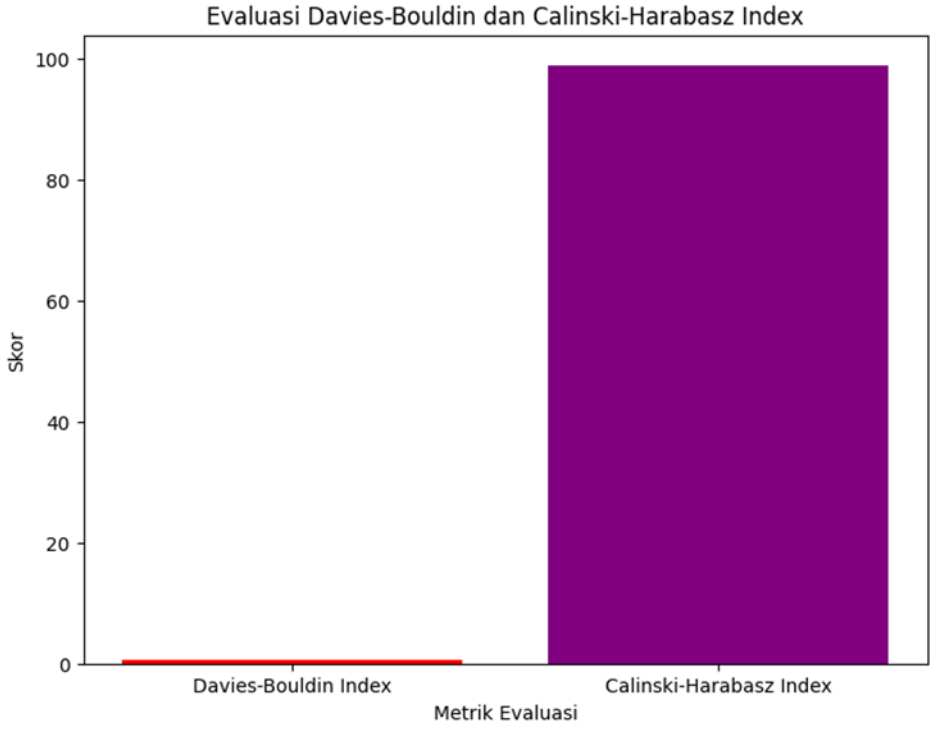
Evaluasi kualitas klasterisasi dilakukan menggunakan Silhouette Score dan Davies-Bouldin Index. Silhouette Score yang mencapai nilai tertinggi saat data dikelompokkan ke dalam dua klaster memperlihatkan bahwa dua klaster ini memisahkan data dengan cukup baik. Selain itu, Davies-Bouldin Index yang digunakan untuk mengukur kompaksi dan pemisahan antar klaster juga memperkuat hasil bahwa pembagian data menjadi dua klaster adalah yang paling optimal.

Tabel 3 Evaluasi Kualitasi klasterisasi

Silhouette Score (RBF)	Silhouette Score (Nearest Neighbors)	Silhouette Score (PCA)	Davies-Bouldin Score	Calinski-Harabasz Score
0.6677246751455114	0.1960207305823995	0.7965958832692485	0.601816506189001	98.8461915122702



Gambar 5 Silhoutte Score



Gambar 6 Davies-Boulding dan Calinski-Harabaz Indeks

Penelitian ini mengonfirmasi bahwa Spectral Clustering adalah metode yang sangat efektif untuk mengelompokkan properti di Jakarta Selatan berdasarkan harga dan luas tanah. Klaster pertama mencakup properti dengan harga moderat, sedangkan klaster kedua terdiri dari properti mewah yang lebih eksklusif. Hal ini memberikan wawasan penting bagi pengembang properti dan pengambil kebijakan dalam memahami segmentasi pasar dan perbedaan karakteristik antar properti di Jakarta Selatan.

4. KESIMPULAN

Penelitian ini menunjukkan pentingnya metode Spectral Clustering dalam analisis harga properti, khususnya di wilayah Jakarta Selatan. Dengan memanfaatkan teknik ini, penelitian berhasil mengelompokkan properti berdasarkan karakteristik seperti harga, luas tanah, dan luas bangunan. Spectral Clustering unggul dalam menangani data yang kompleks dan non-linear, di mana metode konvensional seperti K-Means sering kali kurang efektif.

Spectral Clustering telah berkembang menjadi solusi yang efektif untuk menangani data yang kompleks. Metode ini unggul dalam mengidentifikasi pola non-konveks, yang sangat berguna dalam analisis harga rumah, di mana faktor geografis dan sosio-ekonomi sering membentuk klaster yang tidak teratur. Spectral Clustering menggabungkan prinsip-prinsip teori graf dan dekomposisi spektral untuk memproyeksikan data ke ruang dimensi lebih rendah, sehingga mempermudah proses pengelompokan.

Hasil penelitian menunjukkan bahwa mayoritas properti berada di klaster yang mewakili rumah dengan harga dan luas tanah yang moderat, sedangkan properti dengan harga dan luas yang lebih ekstrem membentuk klaster terpisah. Evaluasi menggunakan Silhouette Score, Davies-Bouldin Index, dan Calinski-Harabasz Score memperkuat bahwa dua hingga tiga klaster adalah jumlah yang optimal dalam pembagian data. Silhouette Score menunjukkan pemisahan klaster yang baik dengan nilai yang lebih tinggi, Davies-Bouldin Index menunjukkan kompaksi dan pemisahan antar klaster dengan nilai yang lebih rendah, dan Calinski-Harabasz Score menilai rasio dispersi antar klaster dengan dispersi intra-klaster dengan nilai yang lebih tinggi.

Dapat disimpulkan bahwa, penerapan Spectral Clustering tidak hanya memberikan wawasan yang lebih mendalam mengenai pola harga rumah di Jakarta Selatan, tetapi juga menawarkan solusi yang lebih akurat dalam pengelompokan data pasar properti. Hasil ini sangat bermanfaat bagi pengembang properti, pembuat kebijakan, dan investor dalam mengambil keputusan yang lebih tepat.

5. DAFTAR PUSTAKA

A. K. Jain, "Data clustering: 50 years beyond k-means," Pattern Recognition Letters, vol. 31, no. 8, pp. 651-666, 2010.

A. N. Rais, W. Warjiyono, I. Alfarobi, S. W. Hadi and W. Kurniawan, "Analisa Prediksi Harga Jual Rumah Menggunakan Algoritma Random Forest Machine Learning," Jurnal Riset Sistem Informasi dan Teknologi Informasi (JURSISTEKNI), vol. 6, no. 2, pp. 416-423, 2024.

A. Yusuf and H. Tjandrasa, "Prediksi Nilai Dengan Metode Spectral Clustering dan Clusterwise Regression," Jurnal SimanteC, vol. 4, no. 1, pp. 1-8, 2014.

B. G. Aji, D. C. Aditya, M. R. Gifari and S. Wijayanto, "Penerapan Algoritma K-Means Untuk Clustering Harga Rumah Di Bandung," Jurnal Ilmiah Informatika Global (UIGM), vol. 14, no. 2, pp. 17-23, 2023.

D. A. Tarigan, "Optimization of the K-Means Clustering Algorithm Using Davies Bouldin Index in Iris Data Classification," International Journal of Data Science, vol. 18, no. 2, pp. 45 - 60, 2024.

D. Chandraderia, V. N. Siwi and S. Fevriera, "Analisis Faktor-Faktor yang Mempengaruhi Harga Rumah di Kawasan Aglomerasi Yogyakarta," Jurnal Pembangunan Wilayah dan Yogyakarta, vol. 18, no. 2, pp. 128-139, Juni 2022.

F. Frisca, A. Bustamam and T. Siswantining, "Implementation Of Spectral Clustering On Microarray Data Of Carcinoma Using K-Means Algorithm," AIP Conference Proceedings, vol. 1825, no. 1, 2017.

M. L. Mutashim, S. A. Damayanti, H. N. Zaki, T. Muhayat and R. Wirawan, "Analisis Prediksi

Harga Rumah Sesuai Spesifikasi Menggunakan Multiple Linear Regression," Jurnal Informatika, vol. 17, no. 3, pp. 238-245, November 2021.
M. Taj, A. E. Mouden, Jakimi and Hajar, "A New Spectral Clustering Approach To Detecting Communities In Graphs," : Proceedings of the 2005 SIAM International Conference on Data Mining, vol. 1, no. 1, pp. 2581-6195, 2018.
M. W. Putri, R. Wasono and I. M. Nur, "Implementasi Spectral Clustering Algortihma Untuk Pengelompokan Sasaran Vaksinasi Covid-19 di Indonesia," Jurnal Statiska, vol. 10, no. 1, pp. 26-31, 2022.
N. Diandra, M. N. Afla and O. Saputra, "Tinjauan Rumah Tinggal Berdasarkan Konsep Rumah Sehatmenurut Regulasi Pemerintah," JTD : Jurnal Ilmiah Teknologi dan Desain Institut Sains dan Teknologi Pradita, pp. 45-44, 2020.
S. Wulandari and D. Novita, "Analisis Clustering Virus MERS-Cov Menggunakan Metode Spectral Clustering Dan Algoritma KMeans," STRING (Satuan Tulisan Riset Dan Inovasi Teknologi), vol. 5, no. 3, pp. 315-323, 2021.
X. Wang and Y. Xu, "An Improved Index For Clustering Validation Based On Silhouette Index and Calinski-Harabasz Index," IOP Conference Series: Materials Science and Engineering, vol. 569, no. 5, p. 052024, 2020.
Y. Hasan, "Pengukuran Silhouette Score dan Davies-Bouldin Index pada Hasil Cluster K-Means dan DBSCAN," KAKIFIKOM (Kumpulan Artikel Karya Ilmiah Fakultas Ilmu Komputer), pp. 60-74, 2024.
Y. Mardi, "Preprocessing Data: Pengumpulan, Pembersihan, Transformasi, dan Penyimpanan," Data Mining : Klasifikasi Menggunakan Algoritma C4.5, vol. 2, no. 3, pp. 213-219, 2017.