

## Analisis Clustering Untuk Segmentasi Wilayah Berdasarkan Karakteristik PBB di Kabupaten Sragen

Febriaji Primadeni<sup>\*1</sup>, Afu Ichsan Pradana<sup>2</sup>, Eko Purwanto<sup>3</sup>

<sup>1,2,3</sup>Teknik Informatika, Fakultas Ilmu Komputer, Universitas Duta Bangsa, Surakarta, Indonesia  
Email: <sup>1</sup>[230103296@mhs.udb.ac.id](mailto:230103296@mhs.udb.ac.id), <sup>2</sup>[afu\\_ichsan@udb.ac.id](mailto:afu_ichsan@udb.ac.id), <sup>3</sup>[eko\\_purwanto@udb.ac.id](mailto:eko_purwanto@udb.ac.id)

### Abstrak

Penelitian ini menganalisis karakteristik Pajak Bumi dan Bangunan (PBB) di Kabupaten Sragen menggunakan metode clustering K-means untuk mengoptimalkan pengelolaan pajak daerah. Analisis dilakukan terhadap dataset yang mencakup 1,5 juta item data PBB periode 2021-2023 dari 196 desa. Variabel yang digunakan meliputi luas bumi, nilai objek pajak, realisasi pembayaran, dan status administrasi pajak. Hasil penelitian menunjukkan pembentukan empat cluster dengan karakteristik berbeda: cluster urban premium (NJOP tinggi, luas lahan kecil), cluster permukiman menengah (NJOP moderat, luas sedang), cluster sub-urban (NJOP rendah, luas besar), dan cluster rural (NJOP terendah, luas terbesar). Analisis korelasi mengungkapkan hubungan positif kuat antara NJOP bangunan dengan PBB ( $r=0,98$ ) dan hubungan negatif antara luas bumi dengan NJOP ( $r=-0,39$ ), mengindikasikan fenomena densifikasi di pusat kabupaten. Pola distribusi desa dalam cluster mengalami perubahan signifikan selama periode penelitian, mencerminkan dinamika produktivitas dan efisiensi pengelolaan PBB. Hasil clustering ini memberikan dasar untuk pengembangan strategi pengelolaan PBB yang lebih terarah dan efektif sesuai karakteristik wilayah.

**Kata kunci:** *clustering k-means, njop daerah, pajak bumi bangunan, pengelolaan pajak daerah, segmentasi wilayah, strategi perpajakan*

## Clustering Analysis for Regional Segmentation Based on Land and Building Tax Characteristics in Sragen Regency

### Abstract

This study analyzes the characteristics of Land and Building Tax (PBB) in Sragen Regency using the K-means clustering method to optimize local tax management. The analysis was conducted on a dataset covering 1.5 million PBB data items from 2021-2023 across 196 villages. Variables included land area, tax object value, payment realization, and tax administration status. The results revealed four distinct clusters: premium urban clusters (high tax object value, small land area), middle-residential clusters (moderate tax object value, medium area), sub-urban clusters (low tax object value, large area), and rural clusters (lowest tax object value, largest area). Correlation analysis showed a strong positive relationship between building tax object value and PBB ( $r=0.98$ ) and a negative relationship between land area and tax object value ( $r=-0.39$ ), indicating densification phenomena in the regency center. Village distribution patterns within clusters underwent significant changes during the study period, reflecting the dynamics of PBB management productivity and efficiency. These clustering results provide a foundation for developing more targeted and effective PBB management strategies according to regional characteristics.

**Keywords:** *k-means clustering, land and building tax, local tax management, regional segmentation, regional tax value, taxation strategy*

## 1. PENDAHULUAN

Pajak Bumi dan Bangunan (PBB) merupakan salah satu instrumen fiskal strategis yang berkontribusi signifikan terhadap Pendapatan Asli Daerah (PAD) dalam mendukung pembangunan infrastruktur dan pelayanan publik di era otonomi daerah [1] [2]. Implementasi desentralisasi fiskal telah memberikan kewenangan lebih luas kepada pemerintah daerah dalam mengelola sumber-sumber pendapatan, termasuk optimalisasi PBB sebagai komponen vital dalam memperkuat kemandirian keuangan daerah[3].

Kabupaten Sragen mencatatkan capaian positif dalam pengelolaan PBB dengan realisasi pendapatan sebesar Rp38 miliar pada tahun 2023, melampaui target yang ditetapkan sebesar Rp35,5 miliar. Namun, di balik

pencapaian tersebut, terdapat permasalahan fundamental yang perlu mendapat perhatian. Data Badan Pengelolaan Keuangan dan Pendapatan Daerah (BPKPD) Kabupaten Sragen mengindikasikan bahwa hingga pertengahan 2023, dari total 196 desa yang ada, hanya 28,57% atau 56 desa yang mencapai pelunasan PBB secara penuh. Kondisi ini mencerminkan adanya tantangan signifikan dalam aspek kepatuhan wajib pajak dan efektivitas sistem penagihan.

Meskipun berbagai inovasi telah diimplementasikan, seperti penerapan sistem penagihan berbasis robot virtual (*chatbot*), kendala literasi teknologi di kalangan masyarakat masih menjadi hambatan utama dalam optimalisasi penerimaan PBB[4]. Situasi ini membutuhkan pendekatan analitis yang lebih komprehensif untuk mengidentifikasi pola kepatuhan wajib pajak [5] dan potensi penerimaan berdasarkan karakteristik wilayah.

Penelitian terdahulu dalam konteks PBB cenderung berfokus pada analisis pajak terhutang [6] atau studi kasus pada lingkup kelurahan [7]. Berbeda dengan penelitian sebelumnya, studi ini mengambil perspektif yang lebih luas dengan menganalisis data pada tingkat kabupaten menggunakan pendekatan *clustering* untuk segmentasi wilayah. Pemilihan metode *clustering* didasarkan pada kemampuannya dalam menghasilkan interpretasi yang lebih mudah dipahami dibandingkan metode fuzzy atau machine learning lanjut [8], terutama dalam mengolah dataset komprehensif yang mencakup 1,5 juta item selama periode 2021-2023.

Penelitian ini bertujuan mengimplementasikan dan mengevaluasi efektivitas metode *clustering* dalam mengidentifikasi pola kepatuhan wajib pajak serta potensi penerimaan PBB berdasarkan karakteristik wilayah di Kabupaten Sragen untuk menghasilkan rekomendasi strategis bagi optimalisasi pengelolaan pajak daerah. Hasil penelitian diharapkan dapat memberikan kontribusi teoretis dalam pengembangan model analisis perpajakan daerah berbasis *clustering*, serta manfaat praktis bagi BPKPD Kabupaten Sragen dalam merumuskan strategi peningkatan efektivitas pengelolaan PBB.

## 2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *clustering analysis* untuk mengoptimalkan pengelolaan Pajak Bumi dan Bangunan (PBB) di Kabupaten Sragen. Metode K-means clustering digunakan sebagai teknik analisis utama karena mampu mengelompokkan data berdasarkan karakteristik yang serupa [9], sehingga efektif dalam mengidentifikasi pola dan distribusi pembayaran PBB. Penggunaan data time series periode 2021-2023 memungkinkan analisis yang komprehensif terhadap tren dan pola pembayaran PBB. Kombinasi antara analisis statistik dan visualisasi data memberikan pemahaman mendalam tentang karakteristik setiap kelompok wajib pajak[10], yang selanjutnya menjadi dasar pengembangan strategi pengelolaan PBB yang lebih efektif.

Tahap awal penelitian dimulai dengan (1) persiapan dan pemrosesan data PBB Kabupaten Sragen yang dilaksanakan melalui serangkaian tahapan sistematis, meliputi pengumpulan dan verifikasi data dari seluruh desa/kelurahan selama periode 2021-2023. Proses ini mengimplementasikan metodologi komprehensif dalam identifikasi sumber data, validasi dokumen, serta ekstraksi variabel-variabel kritis seperti luas bumi, nilai objek pajak, realisasi pembayaran, dan status administrasi perpajakan. Validasi data dilakukan secara menyeluruh untuk memastikan tingkat akurasi dan representativitas dataset yang akan digunakan dalam tahap preprocessing selanjutnya [11]. Selanjutnya, (2) implementasi metode K-means Clustering diaplikasikan untuk melakukan segmentasi dataset berdasarkan karakteristik yang telah ditentukan. Standardisasi variabel menggunakan fungsi `scale()` diterapkan untuk menjamin keseragaman skala pengukuran. Penetapan jumlah cluster optimal menggunakan Elbow Method [12] dengan evaluasi Within Sum of Squares (WSS) pada rentang nilai  $k=1$  hingga  $k=10$ . Proses iteratif dilaksanakan dengan tujuan optimalisasi perbedaan antarcluster dan minimalisasi variasi intracluster.

Penelitian dilanjutkan dengan (3) analisis karakteristik cluster melalui implementasi fungsi agregasi untuk mengidentifikasi parameter-parameter signifikan seperti rata-rata luas bumi, nilai objek pajak, dan tingkat realisasi pembayaran PBB [13]. Hasil analisis dielaborasi secara kuantitatif untuk menghasilkan deskripsi komprehensif mengenai karakteristik setiap cluster, memberikan pemahaman mendalam tentang pola distribusi dan diferensiasi antarcluster yang terbentuk. Kemudian dilakukan (4) evaluasi pola dan tren pembayaran PBB dengan mengaplikasikan serangkaian uji statistik, meliputi analisis korelasi dan uji varians (ANOVA) untuk mengidentifikasi signifikansi hubungan antarvariabel dan perbedaan antarcluster. Implementasi visualisasi data melalui box plot dan scatter plot digunakan sebagai instrumen pendukung dalam interpretasi hasil analisis statistik.

Tahap akhir penelitian mencakup (5) visualisasi hasil clustering yang diimplementasikan menggunakan library `fviz_cluster()` untuk representasi spasial kelompok data. Rangkaian visualisasi mencakup elbow plot untuk penentuan jumlah cluster optimal, scatter plot untuk representasi distribusi data, dan box plot untuk analisis statistik deskriptif. Berdasarkan hasil clustering, (6) interpretasi komprehensif dilakukan untuk menghasilkan rekomendasi kebijakan perpajakan daerah yang spesifik dan terukur. Setiap cluster dianalisis

secara mendalam untuk mengembangkan strategi yang sesuai dengan karakteristiknya, terutama dalam aspek peningkatan kepatuhan pajak dan efisiensi pengelolaan PBB. Akhirnya, (7) optimalisasi pengelolaan PBB dirancang dengan pendekatan berbasis cluster, menghasilkan strategi pengelolaan yang spesifik dan terarah. Rekomendasi yang dihasilkan mencakup kebijakan-kebijakan strategis untuk meningkatkan kepatuhan pajak, menjamin keadilan fiskal, dan meningkatkan efisiensi pemungutan, dengan tujuan utama mendukung pembangunan daerah berbasis data di Kabupaten Sragen.

### 3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan metode *Clustering K-means* untuk menganalisis karakteristik pembayaran Pajak Bumi dan Bangunan (PBB) di tingkat desa. Penelitian dilakukan dengan menggunakan perangkat lunak R Studio, yang dinilai memiliki kemampuan yang memadai dalam mengolah dataset berskala besar [14], termasuk lebih dari 1,5 juta data PBB Kabupaten Sragen untuk periode 2021–2023. Dataset awal sebelum dilakukan proses *Clustering* terdiri dari 15 variabel, yaitu: NOP, Kode Kecamatan, Kecamatan, Kode Desa/Kelurahan, Desa/Kelurahan, Jalan OP, Nama WP SPPT, Tahun Pajak SPPT, Luas Bumi SPPT, Luas Bangunan SPPT, NJOP Bumi SPPT, NJOP Bangunan SPPT, PBB yang Harus Dibayar SPPT, Status Pembayaran SPPT, dan Jenis Bumi.

Tahap selanjutnya dalam penelitian ini adalah melakukan proses seleksi variabel yang bertujuan untuk mengidentifikasi dan menentukan variabel-variabel yang memiliki relevansi signifikan dengan implementasi *clustering data mining* [15]. Berdasarkan hasil analisis komprehensif terhadap ketersediaan data dan kebutuhan penelitian, ditetapkan sepuluh variabel kunci yang akan digunakan dalam proses analisis, yaitu: Nomor Objek Pajak (NOP), Kecamatan, Kode Desa, Desa/Kelurahan, Luas Bumi, Luas Bangunan, Nilai Jual Objek Pajak (NJOP) Bumi, NJOP Bangunan, PBB yang Dibayar, dan Status Pembayaran. Dalam konteks ini, variabel Kode Desa tetap dipertahankan sebagai salah satu parameter analisis dengan pertimbangan adanya potensi ambiguitas dalam identifikasi wilayah, mengingat terdapat beberapa desa yang memiliki nama identik namun berada pada kecamatan yang berbeda. Hasil akhir dari proses pemilihan variabel tersebut disajikan secara rinci dalam (Tabel 1), yang dapat dilihat pada tabel di bawah.

Tabel 1. Hasil Final Pemilihan Variabel Clustering

NOP	KECAM ATAN	KODE_DES A_KEL	DESA_ KEL	LUAS_B UMI	LUAS_ BNG	NJOP_B UMI	NJOP_ BNG	PBB_BA YAR	STATUS_B AYAR
33.14.150.011.0 16.0044.0	Sumberla wang	33.14.15.2011	Ngargot irto	5594	0	1118800 00	0	54374	SUDAH BAYAR
33.14.060.001.0 09.0119.0	Gondang	33.14.06.2001	Srimuly o	4185	0	8370000 0	0	40678	SUDAH BAYAR
33.14.150.011.0 16.0045.0	Sumberla wang	33.14.15.2011	Ngargot irto	4060	0	8120000 0	0	39463	SUDAH BAYAR
33.14.060.001.0 10.0086.0	Gondang	33.14.06.2001	Srimuly o	1905	0	3810000 0	0	18517	BELUM BAYAR
33.14.120.016.0 13.0071.0	Tanon	33.14.12.2016	Gading	1695	0	2373000 0	0	23730	SUDAH BAYAR
33.14.120.016.0 13.0068.0	Tanon	33.14.12.2016	Gading	992	0	1388800 0	0	5278	SUDAH BAYAR
33.14.100.005.0 08.0077.0	Sragen	33.14.10.1005	Nglorog	179	0	1099060 00	0	44893	SUDAH BAYAR
33.14.080.002.0 13.0143.0	Ngrampal	33.14.08.2002	Bener	330	54	4224000 0	272700 00	47366	BELUM BAYAR
33.14.090.006.0 01.0119.0	Karangma lang	33.14.09.2006	Puro	417	72	5337600 0	363600 00	46042	SUDAH BAYAR
33.14.200.002.0 06.0003.0	Jenar	33.14.20.2002	Ngeprin gan	7000	0	5005000 0	0	29610	SUDAH BAYAR
33.14.160.004.0 14.0188.0	Mondokan	33.14.16.2004	Pare	1547	96	1106105 0	411840 00	36428	SUDAH BAYAR
..... 33.14.010.009.0 05.0251.0	..... Kalijambe	..... 33.14.01.2009	..... Banaran	..... 2000	..... 0	..... 4000000 00	..... 0	..... 182400	..... SUDAH BAYAR

*Preprocessing* data merupakan langkah penting dalam memastikan kualitas data sebelum dilakukan analisis *clustering* [16]. Proses ini diawali dengan instalasi pustaka yang relevan, seperti *tidyverse*, *cluster*, *factoextra*, dan *data.table*, yang mendukung berbagai fungsi untuk pengolahan data. Setelah data diimpor, langkah pertama adalah menangani *missing values* dengan memfilter data sehingga hanya entri lengkap yang dianalisis. Selanjutnya, deteksi dan penanganan *outliers* dilakukan untuk memastikan distribusi data yang lebih representatif. Variabel kategorik, seperti STATUS\_BAYAR, dikonversi ke format numerik untuk

memungkinkan analisis statistik lebih lanjut. Data kemudian difilter pada tingkat desa/kelurahan melalui proses agregasi menggunakan fungsi `group_by()` dan `summarise()` dari paket `tidyverse`, dengan setiap variabel numerik, seperti `LUAS BUMI`, `NJOP BUMI`, `NJOP BNG`, dan `PBB BAYAR`, diagregasi menggunakan rata-rata aritmatika. Selain itu, `STATUS BAYAR` dihitung sebagai persentase pembayaran. Pada tahap akhir preprocessing, data distandardisasi menggunakan metode *z-score normalization* untuk mengeliminasi perbedaan skala pengukuran antar variabel, dengan formula

$$z = (x - \mu) / \sigma \quad (1)$$

Keterangan:

$x$  adalah nilai asli

$\mu$  adalah mean

$\sigma$  adalah standar deviasi diimplementasikan dalam kode: `scale(Cluster_data)`

Penjelasan: Mentransformasi data ke skala standar untuk memastikan semua variabel memiliki kontribusi yang setara.

### 3.1. Implementasi *K-means Clustering*

Implementasi *K-means Clustering* dilakukan menggunakan fungsi `kmeans()` dari *R base package* dengan parameter `nstart = 25` untuk *multiple random initialization*, yang bertujuan meminimalkan risiko konvergensi pada *local optima*. Standardisasi data menggunakan *z-score transformation* melalui fungsi `scale()` diterapkan untuk menormalisasi empat variabel utama (luas bumi, NJOP bumi, NJOP bangunan, dan nilai PBB) ke dalam skala yang setara dengan formula  $z = (x - \mu) / \sigma$ , dimana  $x$  adalah nilai observasi,  $\mu$  adalah mean, dan  $\sigma$  adalah standar deviasi. Proses *Clustering* dieksekusi dengan kode `optimal_kmeans <- kmeans(scaled_data, centers = optimal_k, nstart = 25)`. Fungsi `kmeans()` secara internal mengimplementasikan algoritma Hartigan-Wong yang melakukan iterasi dua tahap utama: (1) Assignment Step, yang mengalokasikan setiap observasi ke *Cluster* terdekat berdasarkan jarak Euclidean minimum, dan (2) Update Step, yang memperbarui posisi centroid berdasarkan rata-rata arithmetik dari observasi dalam setiap *Cluster*.

Konvergensi algoritma tercapai ketika tidak ada perubahan dalam assignment observasi ke *Cluster* atau perubahan posisi centroid antar iterasi di bawah threshold yang ditentukan. Visualisasi hasil *Clustering* menggunakan package `factoextra` melalui fungsi `fviz_Cluster()` yang menghasilkan plot dua dimensi dengan Principal Component Analysis (PCA) untuk mereduksi dimensi data dari 4D ke 2D. PCA menghasilkan kombinasi linear dari variabel asli yang memaksimalkan variansi data yang dapat dijelaskan dalam dimensi yang lebih rendah, memungkinkan visualisasi dan interpretasi yang lebih mudah dari hasil *Clustering*. Validasi hasil *Clustering* dilakukan melalui analisis *Within-Cluster Sum of Squares (WSS)* yang diminimalisasi melalui *multiple random initialization*.

Validasi hasil *clustering* dilakukan dengan menganalisis total *within-Cluster sum of squares (WSS)* yang diminimalisasi melalui *multiple random initialization*. Metode Elbow digunakan untuk menentukan jumlah *cluster* optimal dengan membandingkan nilai WSS pada berbagai nilai  $k$  [17]. Penentuan jumlah *Cluster* optimal menggunakan metode Elbow, dengan menganalisis total *Within-Cluster Sum of Squares (WSS)* untuk  $k = 1$  hingga 10 *Cluster*. WSS dihitung menggunakan formula:

$$WSS = \sum_i \sum_j \|x_j - \mu_i\|^2 \quad (2)$$

Dimana:

$x_j$  adalah data point

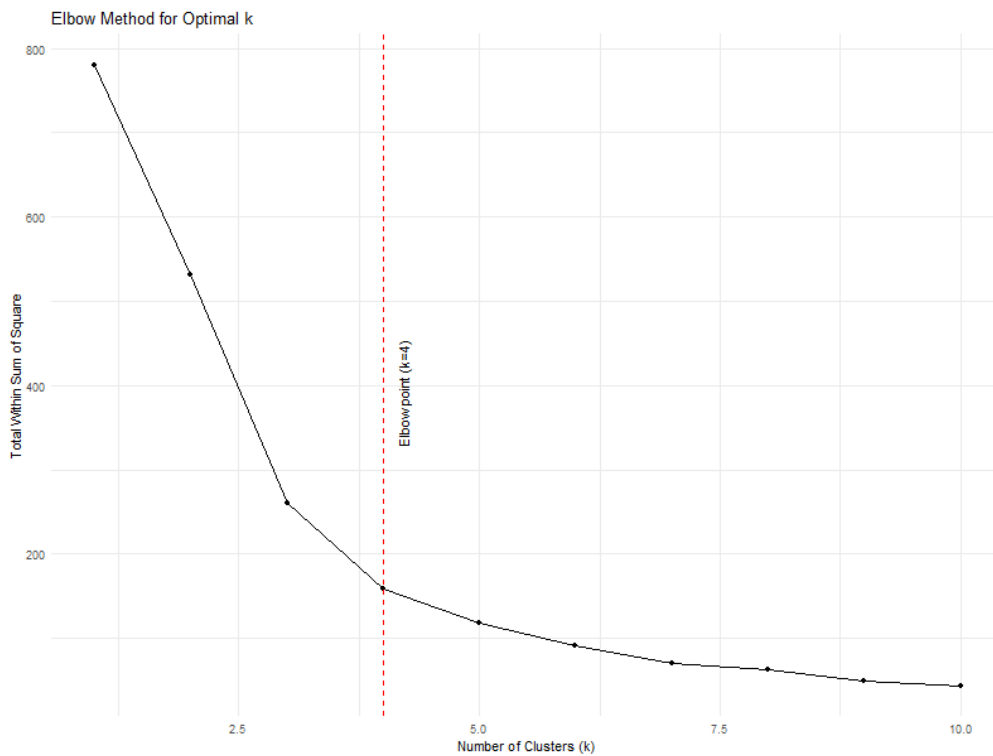
$\mu_i$  adalah centroid *Cluster*  $i$

$\|.\|$  adalah jarak Euclidean Diimplementasikan dalam: `km$tot.withinss`

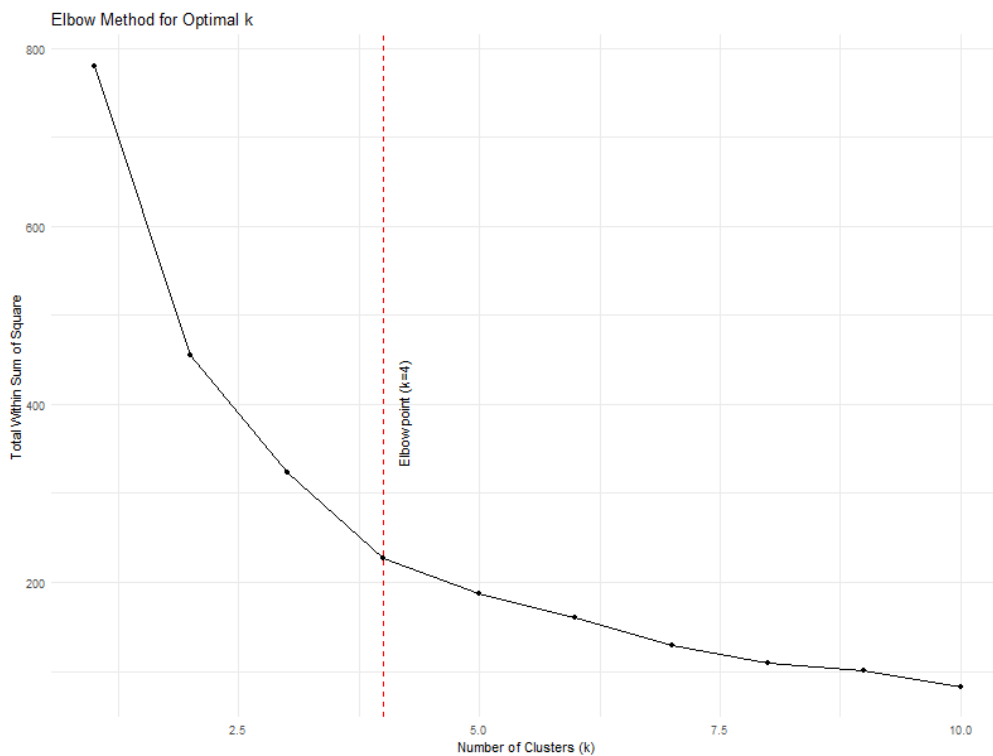
Penjelasan: Mengukur variasi total dalam *cluster* dengan menjumlahkan jarak kuadrat antara setiap titik data dan centroid klusternya.

Hasil analisis menggunakan metode Elbow (Elbow Method) pada dataset Pajak Bumi dan Bangunan (PBB) Kabupaten Sragen selama periode tahun 2021 hingga 2023 menunjukkan bahwa jumlah *cluster* optimal adalah empat. Hal ini ditunjukkan pada Gambar 1, Gambar 2, dan Gambar 3, di mana titik siku (elbow point) terlihat jelas pada  $k = 4$ , yang mengindikasikan bahwa penambahan jumlah *cluster* setelah nilai ini tidak memberikan peningkatan signifikan terhadap kualitas pengelompokan data [18]. Hal ini didasarkan pada observasi terhadap pola penurunan nilai *Within Sum of Squares (WSS)*, di mana terlihat adanya titik siku (elbow point) yang signifikan pada  $k = 4$ , yang ditandai dengan mulai melandainya kurva penurunan WSS setelah titik tersebut.

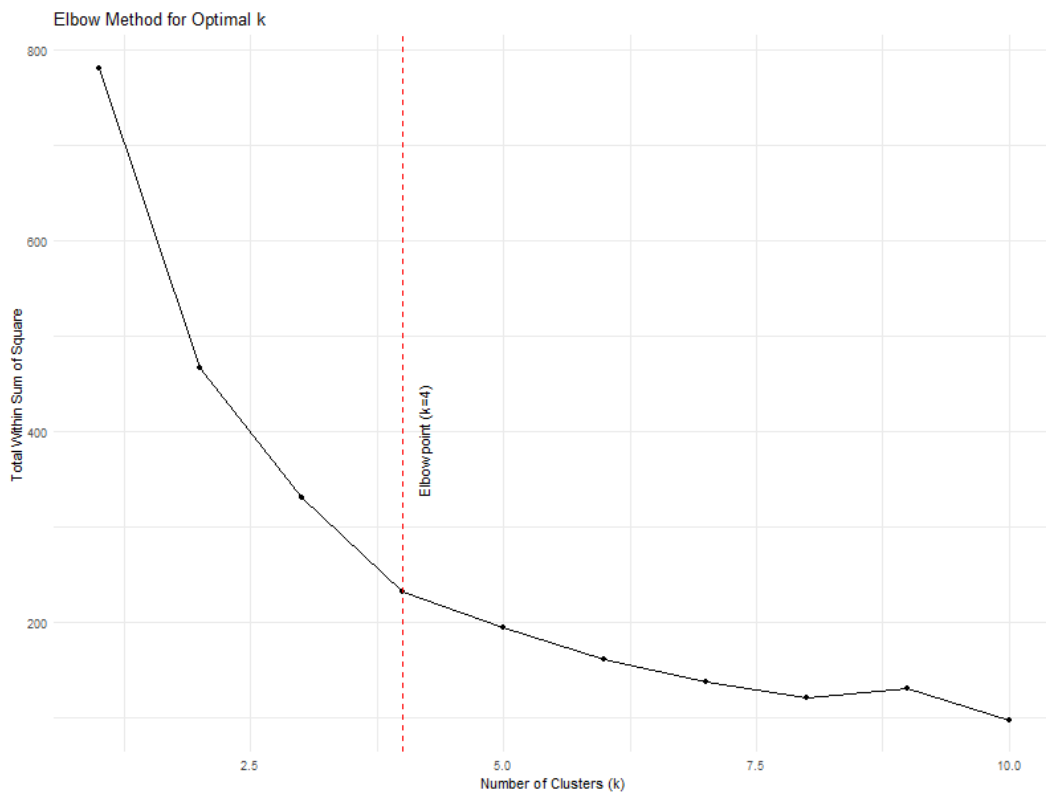
Penurunan WSS yang tidak terlalu signifikan setelah  $k = 4$  mengindikasikan bahwa penambahan jumlah cluster lebih lanjut tidak memberikan peningkatan yang substansial terhadap kualitas pengelompokan data, sehingga dapat disimpulkan bahwa penggunaan 4 cluster merupakan solusi yang optimal untuk segmentasi dataset PBB pada wilayah tersebut.



Gambar 1. Grafik Elbow Method untuk Periode 2021



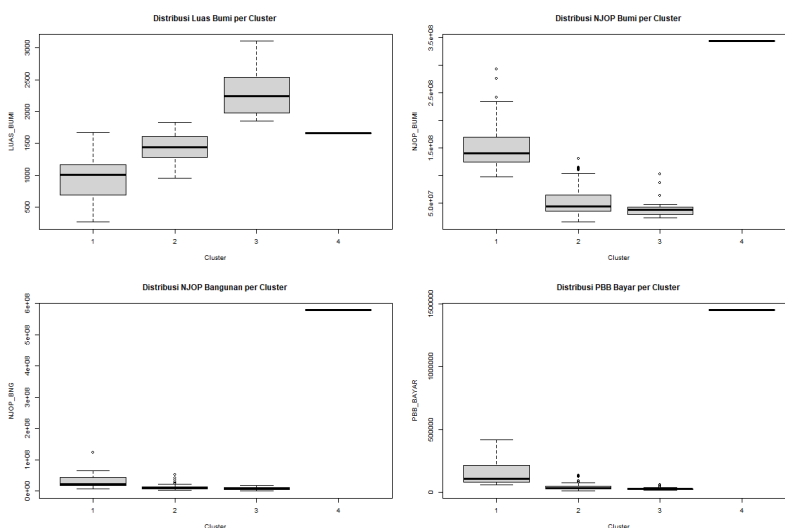
Gambar 2. Grafik Elbow Method untuk Periode 2022



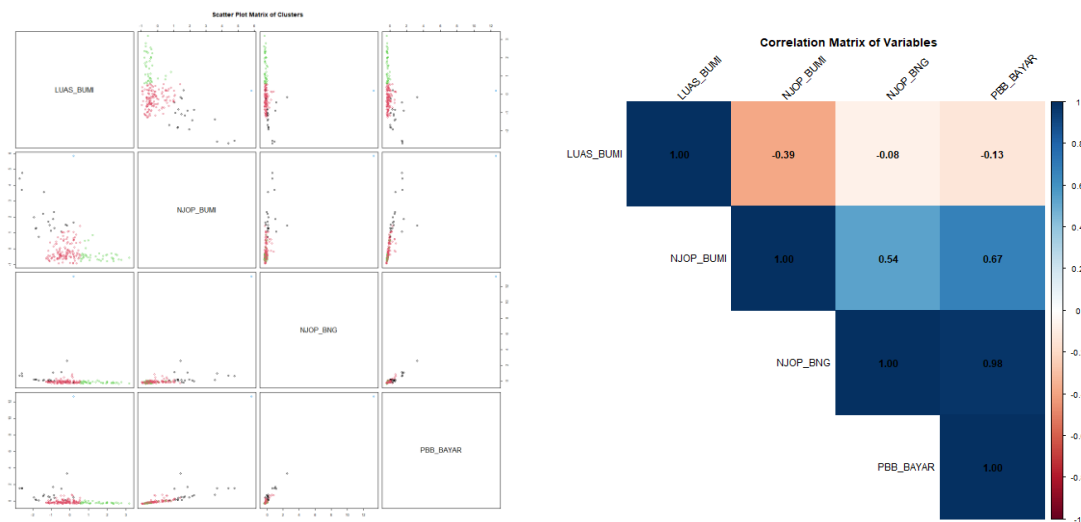
Gambar 3. Grafik Elbow Method untuk Periode 2023

### 3.2. Hasil Analisis PBB Kabupaten Sragen Tahun 2021-2023

Distribusi cluster berdasarkan karakteristik PBB untuk tahun 2021 yang ditampilkan *cluster bloxplots* pada Gambar 4, analisis *clustering* menggunakan visualisasi boxplot mengidentifikasi empat karakteristik wilayah yang berbeda di Kabupaten Sragen. *Cluster 1* dicirikan dengan luas bumi median 1000m<sup>2</sup> yang merepresentasikan kawasan permukiman kepadatan sedang, disertai nilai NJOP bumi dan bangunan kategori moderat-tinggi. *Cluster 2* menunjukkan median luas bumi lebih besar (1500m<sup>2</sup>) dengan beberapa outlier di atas 2000m<sup>2</sup>, namun memiliki NJOP lebih rendah, mencerminkan karakteristik kawasan sub-urban dengan nilai properti terjangkau. *Cluster 3* memiliki median luas bumi tertinggi (2500m<sup>2</sup>) dengan NJOP terendah, mengindikasikan karakteristik kawasan rural atau pertanian. Sementara itu, *cluster 4* menampilkan distribusi luas yang stabil (median 1600m<sup>2</sup>) dengan NJOP tertinggi, menandakan kawasan premium.



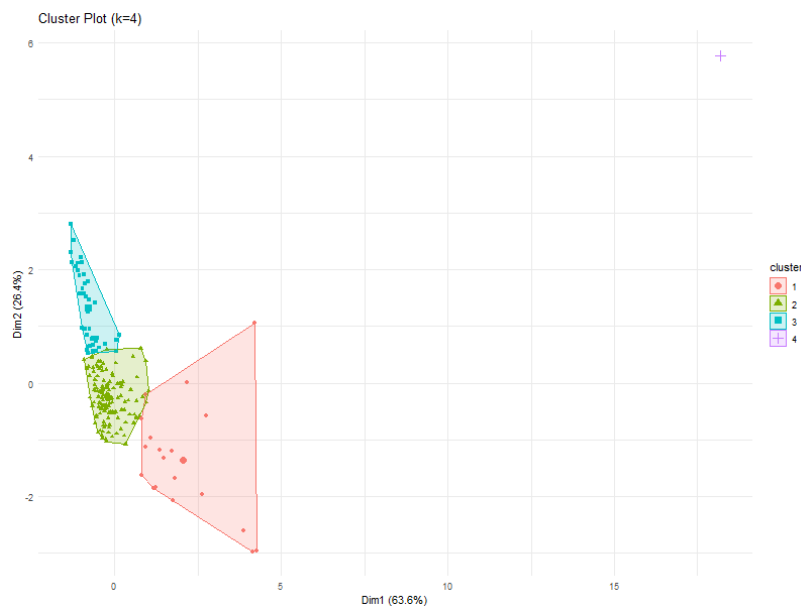
Gambar 4. Cluster Boxplots PBB Tahun 2021



Gambar 5. Scatter Plot Matrix dan Matrix Korelasi PBB Tahun 2021

Analisis pada Gambar 5 menampilkan scatter plot matrix dan matrix korelasi tahun 2021 mengungkapkan kompleksitas hubungan antar variabel, di mana relasi antara luas bumi dengan NJOP bumi menunjukkan pola sebaran acak yang mengindikasikan heterogenitas penggunaan lahan. Terdapat *clustering* yang jelas antara luas bumi dengan NJOP bangunan yang mencerminkan zonasi pembangunan terencana. Hasil analisis korelasi menunjukkan hubungan positif kuat antara NJOP bangunan dengan PBB bayar ( $r=0,98$ ), mengindikasikan dominasi nilai bangunan dalam penentuan pajak. Korelasi negatif antara luas bumi dengan NJOP bumi ( $r=-0,39$ ) mengonfirmasi fenomena densifikasi di pusat kabupaten.

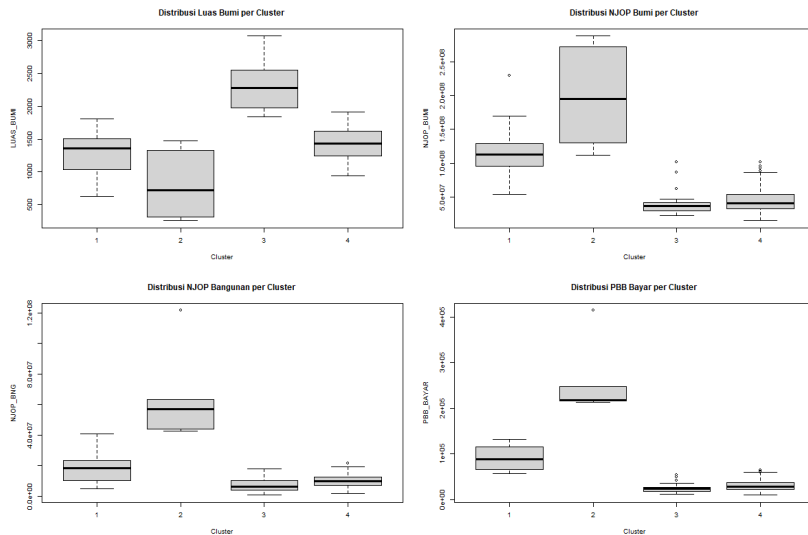
Lebih lanjut, analisis data PBB Kabupaten Sragen tahun 2021 mengungkapkan pola kompleks hubungan antara luas lahan dan nilainya. Menariknya, luas tanah tidak selalu berbanding lurus dengan harga, yang ditunjukkan oleh hubungan negatif lemah ( $r=-0,39$ ). Kondisi ini menjelaskan bahwa faktor-faktor seperti lokasi, infrastruktur, dan potensi pengembangan wilayah memainkan peran penting dalam menentukan nilai properti. Artinya, sebidang tanah yang lebih kecil di lokasi strategis bisa jadi memiliki nilai jauh lebih tinggi dibandingkan lahan yang lebih luas di area kurang berkembang.



Gambar 6. Plot Klaster (k=4) Optimal Tahun 2021

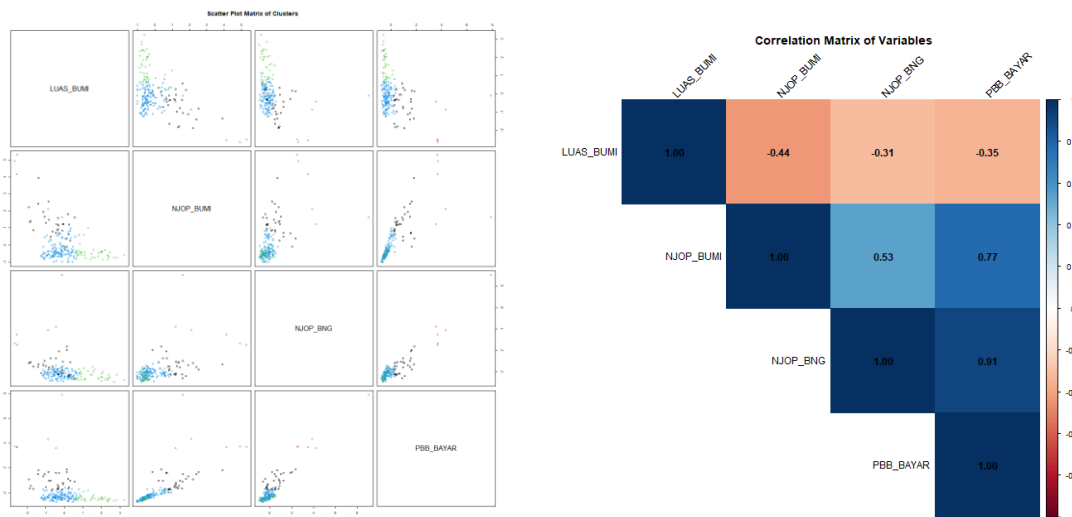
Sebagaimana terlihat pada Gambar 6, visualisasi *cluster* plot dengan  $k=4$  pada tahun 2021 memberikan gambaran struktur spasial Kabupaten Sragen, dengan Dim1 menjelaskan 63,6% variasi berdasarkan nilai ekonomis properti. *Cluster* 1 merepresentasikan wilayah urban menengah, *Cluster* 2 mencerminkan wilayah sub-

urban dengan tingkat pembayaran PBB lebih tinggi, *Cluster 3* menunjukkan wilayah rural dengan luas lahan rata-rata tertinggi, dan *Cluster 4* menunjukkan wilayah urban premium dengan pembayaran PBB rata-rata tertinggi namun tingkat kepatuhan yang lebih rendah.



Gambar 7. Cluster Boxplots PBB Tahun 2022

Distribusi cluster untuk tahun 2022 disajikan dalam Gambar 7, visualisasi boxplot mengidentifikasi karakteristik yang berbeda dari tahun sebelumnya. *Cluster 1* memiliki karakteristik luas bumi dengan median 1250m<sup>2</sup> mencerminkan kawasan permukiman kepadatan menengah dengan NJOP kategori moderat. *Cluster 2* menunjukkan median luas bumi relatif rendah (750m<sup>2</sup>) namun memiliki NJOP tertinggi, mengindikasikan kawasan premium atau pusat kabupaten dengan densitas tinggi. *Cluster 3* memiliki median luas bumi tertinggi (2250m<sup>2</sup>) dengan NJOP moderat-tinggi, merepresentasikan kawasan pengembangan baru. *Cluster 4* menampilkan distribusi luas stabil (median 1500m<sup>2</sup>) dengan NJOP terendah, mengindikasikan kawasan rural.

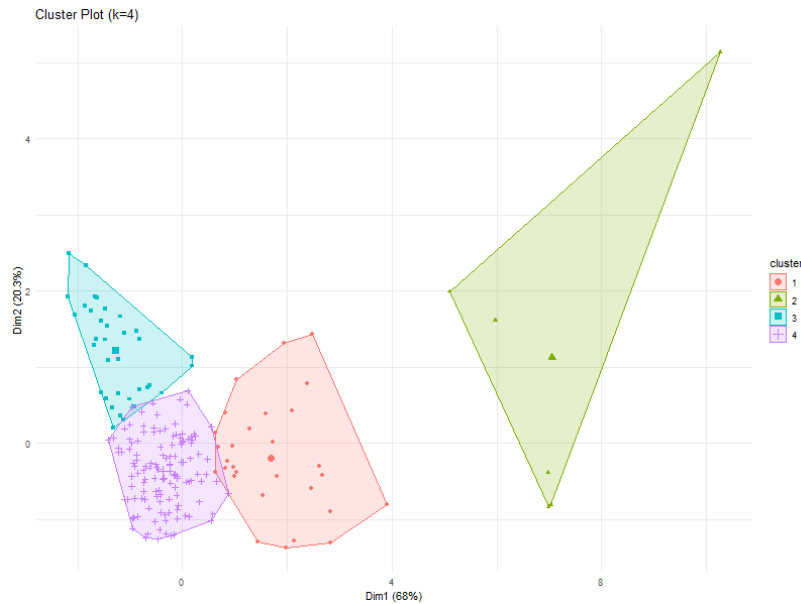


Gambar 8. Scatter Plot Matrix dan Matrix Korelasi PBB Tahun 2022

Analisis scatter plot matrix tahun 2022 menunjukkan pola hubungan yang lebih sistematis dibandingkan tahun sebelumnya. Hubungan variabel PBB pada tahun 2022 divisualisasikan dalam Gambar 8 antara luas bumi dengan NJOP bumi menunjukkan pola *clustering* yang jelas, mengindikasikan zonasi terstruktur dalam penggunaan lahan. Hasil analisis korelasi menunjukkan hubungan negatif moderat antara luas bumi dengan NJOP bumi ( $r=-0,44$ ), mencerminkan fenomena inverse relationship yang lebih kuat. Korelasi positif kuat ditemukan antara NJOP bangunan dengan PBB bayar ( $r=0,91$ ) dan NJOP bumi dengan PBB bayar ( $r=0,77$ ).

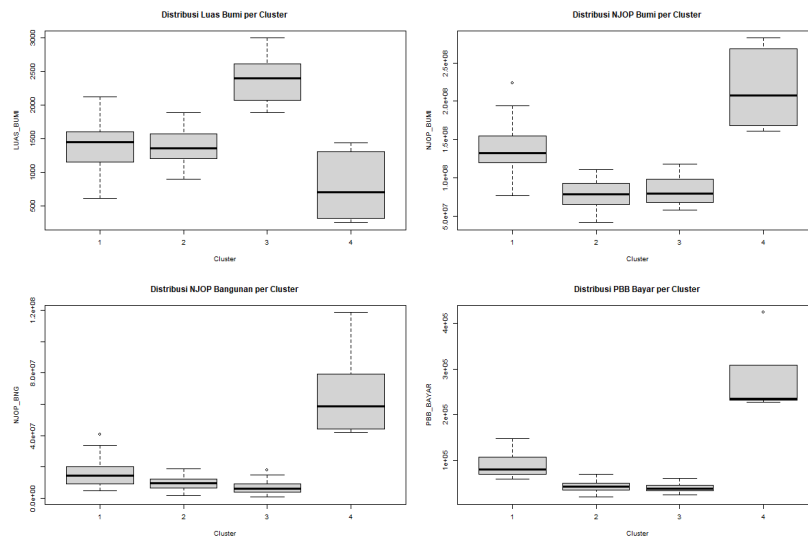


Sejalan dengan analisis tersebut, pada tahun 2022, pola hubungan luas lahan dan nilainya di Kabupaten Sragen mulai menunjukkan struktur yang lebih sistematis. Korelasi negatif moderat ( $r=-0,44$ ) mengindikasikan proses penataan ruang yang semakin terencana. Hal ini menandakan bahwa pemerintah daerah mulai mengoptimalkan penggunaan lahan tidak sekadar berdasarkan luasan, melainkan mempertimbangkan potensi ekonomi dan efisiensi ruang. Hubungan kuat antara nilai bangunan dengan pajak yang dibayarkan ( $r=0,91$ ) menunjukkan bahwa kualitas dan karakteristik bangunan menjadi faktor utama dalam perhitungan pajak properti.



Gambar 9. Plot Kluster (k=4) Optimal Tahun 2022

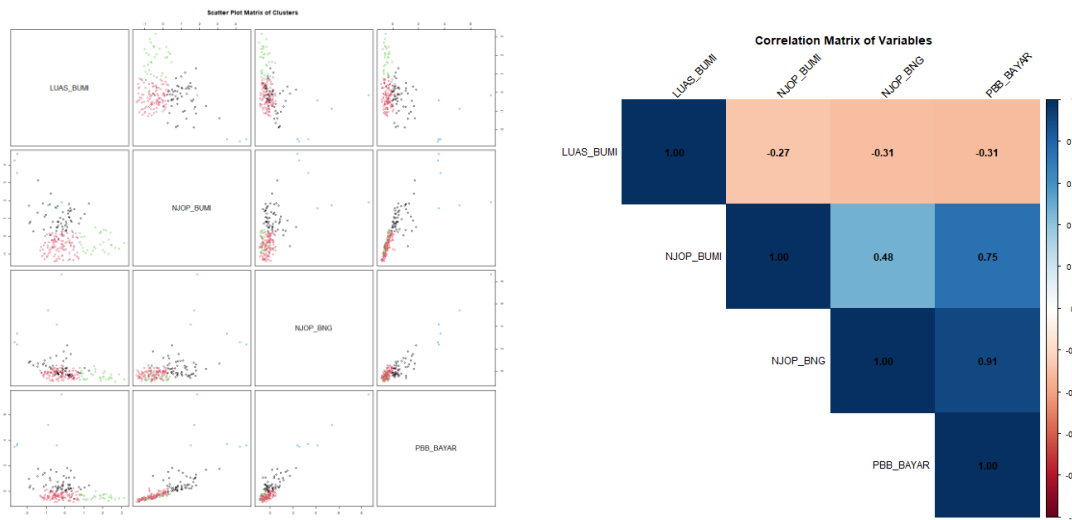
Visualisasi *cluster* plot tahun 2022 seperti terlihat pada Gambar 9, menunjukkan peningkatan signifikan dalam penjelasan variasi oleh Dim1 menjadi 88%, mengindikasikan dominasi faktor ekonomi yang semakin kuat dalam pembentukan *Cluster*. *Cluster* 1 (merah) merepresentasikan kawasan permukiman menengah, *Cluster* 2 (hijau) menandai zona premium, *Cluster* 3 (biru) mengindikasikan kawasan pengembangan baru, dan *Cluster* 4 (ungu) mencirikan kawasan rural.



Gambar 10. Cluster Boxplots PBB Tahun 2023

Pada tahun 2023, visualisasi boxplot menunjukkan perubahan karakteristik yang signifikan. *Cluster* 1 memiliki median luas bumi 1500m<sup>2</sup> dengan variabilitas moderat dan NJOP kategori menengah, mencerminkan kawasan permukiman berkembang. *Cluster* 2 menampilkan median luas bumi 1400m<sup>2</sup> dengan NJOP moderat-

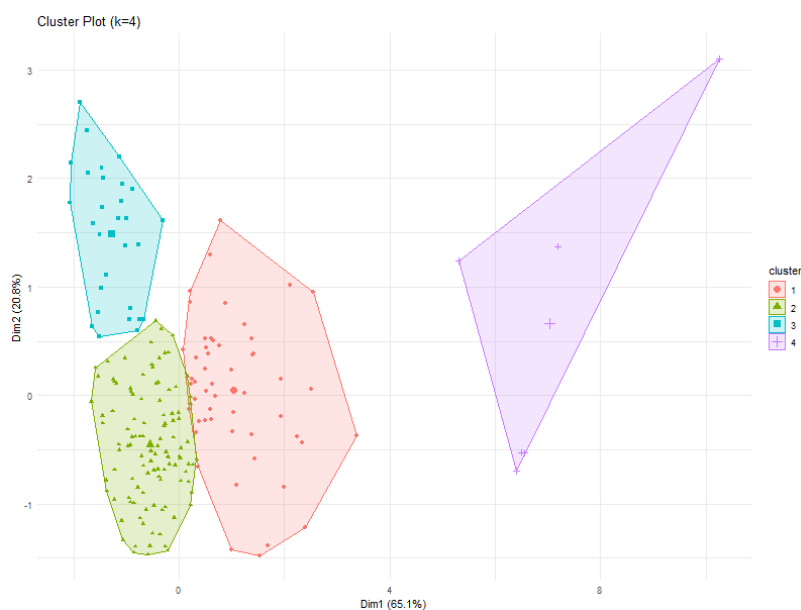
rendah, mengindikasikan kawasan permukiman established. *Cluster 3* memiliki median luas bumi tertinggi (2500m<sup>2</sup>) dengan NJOP moderat-tinggi, merepresentasikan kawasan premium. Karakteristik cluster untuk tahun 2023 ditampilkan dalam Gambar 10, pada *cluster 4* menunjukkan luas bumi relatif kecil (750m<sup>2</sup>) dengan NJOP dan PBB bayar tertinggi, mengindikasikan pusat kabupaten



Gambar 11. Scatter Plot Matrix dan Matrix Korelasi PBB Tahun 2023

Analisis scatter plot matrix tahun 2023 mengungkapkan segregasi yang jelas antar *Cluster* dalam hubungan LUAS BUMI dengan NJOP BUMI, menunjukkan zonasi terstruktur dalam pemanfaatan lahan. Hasil analisis korelasi menunjukkan hubungan negatif moderat yang melemah antara LUAS BUMI dengan NJOP BUMI ( $r = -0,27$ ) dibandingkan tahun sebelumnya. Korelasi positif kuat tetap terjaga antara NJOP\_BNG dengan PBB\_BAYAR ( $r = 0,91$ ) dan NJOP BUMI dengan PBB\_BAYAR ( $r = 0,75$ ).

Sebagaimana divisualisasikan pada Gambar 11, hubungan antara luas bumi dan NJOP bumi melemah menjadi  $r = -0,27$  pada tahun 2023, yang dapat diinterpretasikan sebagai indikasi penyesuaian dan kedewasaan sistem tata ruang daerah. Temuan ini menggambarkan bahwa pengembangan wilayah tidak lagi bergantung secara linear pada luas lahan, melainkan lebih pada kualitas dan potensi pengembangan. Konsistensi hubungan antara nilai bangunan dengan pajak ( $r = 0,91$ ) semakin menegaskan bahwa kualitas bangunan tetap menjadi parameter utama dalam penilaian properti dan penetapan pajak.



Gambar 12. Plot Kluster (k=4) Optimal Tahun 2023

Visualisasi *cluster plot* tahun 2023 dengan Dim1 menjelaskan 65,1% variasi data menunjukkan stabilitas dalam struktur spasial kabupaten. *cluster 4* yang terpisah jauh mencerminkan zona premium dengan nilai ekonomis tinggi. *cluster 1* dan *2* yang berdekatan menunjukkan karakteristik relatif mirip namun tetap dapat dibedakan, sementara *cluster 3* menempati posisi transisi sebagai kawasan pengembangan potensial. Gambar 12 menunjukkan *cluster plot* tahun 2023 dengan *cluster premium* yang semakin terpisah dari *cluster* lainnya, mengindikasikan perbedaan signifikan dalam karakteristik ekonomi. Distribusi data dengan minimal overlap mengkonfirmasi efektivitas pemilihan jumlah *cluster* dalam analisis.

#### 4. TEMUAN SPESIFIK

Analisis *Clustering* terhadap distribusi desa berdasarkan karakteristik ekonomi dan pajak bumi bangunan (PBB) menghasilkan temuan yang signifikan pada periode 2021 hingga 2023. Pada tahun 2021, terdapat empat *Cluster* dengan distribusi jumlah desa yang bervariasi: *Cluster 1* mencakup 19 desa, *Cluster 2* sebanyak 135 desa, *Cluster 3* mencakup 41 desa, dan *Cluster 4* hanya terdiri dari 1 desa. Distribusi ini menunjukkan bahwa *Cluster 2*, yang memiliki disiplin pembayaran pajak tertinggi (97,34%) namun NJOP Bangunan yang relatif rendah, mendominasi secara kuantitas desa. Sebaliknya, *Cluster 4* yang hanya terdiri dari satu desa menunjukkan karakteristik yang unik dengan potensi PBB yang sangat tinggi berkat NJOP Bumi dan Bangunan yang luar biasa.

Pada tahun 2022, pola distribusi mengalami perubahan signifikan. *Cluster 1* berkembang menjadi 29 desa, *Cluster 2* mengecil menjadi hanya 6 desa, *Cluster 3* bertambah menjadi 36 desa, sementara *Cluster 4* mengalami pertumbuhan sangat pesat, mencakup 125 desa. Pergeseran ini menunjukkan adanya redistribusi desa dari *Cluster 2* ke *Cluster 4*, mencerminkan peningkatan NJOP Bangunan dan produktivitas PBB pada sejumlah besar desa. Sementara itu, *Cluster 1* menunjukkan peningkatan jumlah desa dengan karakteristik NJOP yang tetap menengah, sementara *Cluster 3* tetap mendominasi dalam hal luas tanah tetapi tetap mempertahankan potensi PBB yang rendah.

Pada tahun 2023, distribusi desa kembali mengalami perubahan. *Cluster 1* bertambah menjadi 52 desa, *Cluster 2* melonjak menjadi 109 desa, *Cluster 3* menyusut menjadi 29 desa, dan *Cluster 4* tetap mencakup 6 desa. Peningkatan desa dalam *Cluster 1* menunjukkan bahwa lebih banyak desa yang mengalami peningkatan NJOP Bumi tetapi tetap mempertahankan NJOP Bangunan yang rendah. Sementara itu, *Cluster 2*, yang sebelumnya menyusut pada tahun 2022, kembali mendominasi dengan jumlah desa terbesar, mencerminkan peningkatan efisiensi dan keberagaman karakteristik NJOP. Di sisi lain, *Cluster 3* kembali menunjukkan penurunan jumlah desa, mencerminkan adanya perbaikan NJOP pada beberapa desa yang sebelumnya tergolong rendah. *Cluster 4* tetap konsisten dengan jumlah desa yang sedikit namun dengan potensi PBB yang sangat tinggi, menunjukkan bahwa desa-desa ini memiliki efisiensi yang luar biasa dalam memanfaatkan luas tanah yang relatif kecil.

Secara keseluruhan, pola distribusi desa dari tahun ke tahun menunjukkan adanya pergeseran karakteristik yang signifikan pada beberapa *cluster*. *Cluster 1* cenderung mengalami peningkatan jumlah desa secara konsisten, mencerminkan pertumbuhan NJOP Bumi pada desa-desa berpotensi menengah. *Cluster 2*, meskipun mengalami fluktuasi, tetap mencerminkan dominasi secara jumlah desa dan keberagaman karakteristik ekonomi. *Cluster 3* menunjukkan tren menyusut yang mencerminkan adanya peningkatan kualitas NJOP pada beberapa desa, sementara *cluster 4* tetap menjadi *cluster* khusus dengan potensi ekonomi tertinggi meskipun memiliki jumlah desa yang sangat kecil. Perubahan ini menyoroti dinamika produktivitas dan efisiensi desa dalam memanfaatkan potensi PBB dari waktu ke waktu.

#### 5. KESIMPULAN

Analisis *Clustering* terhadap data Pajak Bumi dan Bangunan (PBB) Kabupaten Sragen menggunakan metode *k-means* terbukti efektif dalam mengelompokkan wilayah berdasarkan karakteristik PBB. Dengan mempertimbangkan variabel-variabel seperti luas bumi dan bangunan, nilai jual objek pajak (NJOP) bumi dan bangunan, jumlah pajak yang harus dibayar, serta status pembayaran, *k-means* mampu membentuk beberapa *cluster* yang jelas dan relevan. Hasil *Clustering* ini mengungkapkan pola-pola signifikan, seperti *cluster* dengan NJOP tinggi yang umumnya berada di kawasan perkotaan dengan aktivitas ekonomi yang intensif, sementara *Cluster* dengan NJOP rendah cenderung terletak di daerah perdesaan dengan sektor pertanian sebagai dominasi. Metode ini juga memberikan wawasan strategis tentang potensi pajak di setiap *Cluster*, yang dapat memudahkan pengambilan keputusan dalam penagihan pajak atau kebijakan lain untuk meningkatkan Pendapatan Asli Daerah (PAD). Keefektifan metode *k-means* terlihat dari kemampuannya untuk memvisualisasikan data yang kompleks menjadi segmen-segmen yang mudah dianalisis, sehingga mendukung pemerintah daerah dalam merumuskan kebijakan fiskal yang lebih terarah dan optimal. Dengan memanfaatkan hasil *clustering* ini, strategi penagihan pajak dapat diprioritaskan di wilayah dengan potensi pajak tinggi, sementara pendekatan yang lebih inklusif

dapat diterapkan untuk mendorong kepatuhan pajak di *cluster* lainnya. Hasil analisis ini menjadi landasan yang kuat untuk mendukung pengelolaan PBB yang lebih efisien dan adil, sekaligus mendorong peningkatan PAD secara berkelanjutan.

## DAFTAR PUSTAKA

- [1] M. A. Putri, N. Rahaningsih, F. M. Basysyar, and O. Nurdiawan, "Penerapan Data Mining Menggunakan Metode Clustering Untuk Mengetahui Kelompok Kepatuhan Wajib Pajak Bumi dan Bangunan," *J. Account. Inf. Syst.*, vol. 5, no. 2, pp. 145–156, 2022, doi: 10.32627/aims.v5i2.496.
- [2] O. Daerah and P. Jabar, "Peningkatan pertumbuhan ekonomi provinsi jawa barat berbasis kebijakan desentralisasi fiskal," *PRESTISE*, vol. 4, no. 1, pp. 64–90, 2024.
- [3] E. S. Lodi, S. Alam, and O. Deviany, "Optimizing Rural and Urban Land and Building Tax Receipts in Increasing Local Own Revenue in Enrekang Regency," *J. Gov. Local Polit.*, vol. 5, no. 1, pp. 1–11, 2023.
- [4] E. T. Irianti and F. Niswah, "Optimalisasi Pemungutan Pajak Bumi Dan Bangunan Perdesaan Perkotaan (Pbb-P2) Dalam Meningkatkan Pendapatan Asli Daerah Di Kabupaten Gresik," *Publika*, pp. 503–514, 2021, doi: 10.26740/publika.v9n4.p503-514.
- [5] P. J. Sani and S. Sulfan, "Perilaku Kepatuhan Wajib Pajak Orang Pribadi Pekerjaan Bebas Di Kota Denpasar," *J. Pajak dan Keuang. Negara*, vol. 3, no. 2, pp. 294–304, 2022, doi: 10.31092/jpkn.v3i2.1520.
- [6] J. Mardini, "PENGELOMPOKAN DATA PBB MASYARAKAT," *J. Math. Technol.*, vol. 3, no. May, pp. 54–61, 2024.
- [7] R. Nursaniah, N. Rahaningsih, I. Ali, and N. Dienwati Nuris, "Pengelompokan Data Penerimaan Pajak Bumi Dan Bangunan Berdasarkan Kelurahan Di Kota Tasikmalaya Menggunakan Algoritma K-Means," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, pp. 1477–1483, 2024, doi: 10.36040/jati.v8i2.9009.
- [8] D. M. Putra and F. F. Abdulloh, "Comparison of Clustering Algorithms : Fuzzy C-Means , K-Means , and DBSCAN for House Classification Based on Specifications and Price," *J. Appl. Informatics Comput.*, vol. 8, no. 2, pp. 509–515, 2024.
- [9] M. Adelina Bui and A. Bahtiar, "Implementasi Metode Algoritma K-Means Clustering Untuk Mengelompokkan Transaksi Penjualan Barang Di Toko Arino," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, pp. 1451–1456, 2024, doi: 10.36040/jati.v8i2.8975.
- [10] S. S. Firdaus, Y. Budisusanto, and U. W. Deviantari, "Visualisasi Spasial dan Basis Data Pajak Bumi dan Bangunan (PBB) (Studi Kasus : Desa Bener, Madiun)," *Geoid*, vol. 16, no. 1, p. 131, 2021, doi: 10.12962/j24423998.v16i1.8567.
- [11] A. Agung, A. Daniswara, I. Kadek, and D. Nuryana, "Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru," *J. Informatics Comput. Sci.*, vol. 05, pp. 97–100, 2023.
- [12] N. A. Maori and E. Evanita, "Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 14, no. 2, pp. 277–288, 2023, doi: 10.24176/simet.v14i2.9630.
- [13] S. Mahardi, "PENERAPAN DATA MINING UNTUK CLUSTERING DATA PAJAK BUMI DAN BANGUNAN MENGGUNAKAN ALGORITMA K-MEANS," 2021, *Sekolah Tinggi Manajemen Informatika dan Komputer LIKMI*.
- [14] D. Ariyanto and F. Rachmadiarti, "Peningkatan Kemampuan Analisis Statistik Menggunakan Aplikasi R Studio Berbasis OpenSource Untuk Kebutuhan Penelitian Dosen Di Fakultas Mipa Universitas Negeri Surabaya," *J. Umum Pengabd. Masy.*, pp. 13–20, 2022.
- [15] I. Romli, "Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Klasifikasi Penyakit Ispa," *Indones. J. Bus. Intell.*, vol. 4, no. 1, p. 10, 2021, doi: 10.21927/ijubi.v4i1.1727.
- [16] A. Herlin Lutfiannisa, Maimunah, and P. Sukmasetya, "Clustering Data Pasien Berdasarkan Usia di Puskesmas Menerapkan Metode K-Means," *J. Inf. Syst. Res.*, vol. 5, no. 2, pp. 639–647, 2024, doi: 10.47065/josh.v5i2.4755.
- [17] W. A. Silamantha and K. Hadiono, "Analisis RFM dan K-Means Clustering untuk Segmentasi Pelanggan pada PT . Sanutama Bumi Arto," *KESATRIA J. Penerapan Sist. Inf. (Komputer dan Manajemen)*, vol. 5, no. 3, pp. 1297–1305, 2024.
- [18] I. Wahyudi, M. B. Sulthan, and L. Suhartini, "Analisa Penentuan Cluster Terbaik Pada Metode K-Means

Menggunakan Elbow Terhadap Sentra Industri Produksi Di Pamekasan,” *J. Apl. Teknol. Inf. dan Manaj.*, vol. 2, no. 2, pp. 72–81, 2021, doi: 10.31102/jatim.v2i2.1274.