

Penentuan Prioritas Penagihan Pbb Berbasis K-Means Clustering Di Desa Gunasari

Muhammad Anwar^{1)*}, Esa Firmansyah²⁾, Beben Sutara³⁾

^{1,2,3)} Universitas Sebelas April

Received: 11 December 2025

Accepted: 21 December 2025

Published: 26 December 2025



*220660121121@unsap.student.unsap.ac.id

Kata Kunci: Clustering, Pajak Bangunan Dan Bumi, PBB, Machine Learning

DSI: Jurnal Data Science Indonesia is licensed under a Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0).

Abstrak : Pajak Bumi dan Bangunan(PBB) menjadi salah satu sumber penting pendapatan asli daerah yang berperan dalam membiayai pembangunan di Desa Gunasari. Penelitian ini bertujuan menentukan prioritas operasional penagihan PBB menggunakan K-Means Clustering berbasis data dengan 87 jalan dan total 1.382 tagihan PBB tertunggak tahun 2025. Dua variabel utama digunakan yaitu jumlah tagihan PBB tertunggak per jalan dan total nilai rupiah tertunggak per jalan. Data diperoleh dari sistem administrasi desa pada Januari 2025, kemudian dilakukan feature engineering untuk agregasi level jalan. Elbow Method dan Silhouette Score menunjukkan k=3 sebagai jumlah cluster optimal dengan skor silhouette 0,681 dengan kategori good. Algoritma K-Means membagi 87 jalan menjadi tiga kelompok prioritas yaitu Prioritas Rendah (62 jalan, rata-rata 6,4 tagihan tertunggak, total rata-rata Rp458.544), Prioritas Menengah (21 jalan, rata-rata 42,2 tagihan tertunggak, total rata-rata Rp2.299.109), Prioritas Tinggi (4 jalan, rata-rata 25,2 tagihan tertunggak, total rata-rata Rp7.328.102). Prioritas tinggi didefinisikan berdasarkan total nilai nominal tertunggak yang signifikan secara finansial (value-based), sedangkan prioritas menengah berdasarkan volume tagihan tertunggak yang tinggi (volume-based). Pendekatan ini memungkinkan pemerintah desa mengalokasikan sumber daya penagihan lebih efisien dengan fokus pada jalan prioritas tinggi untuk penagihan strategis dan jalan prioritas menengah untuk sosialisasi intensif.

PENDAHULUAN

Pajak Bumi dan Bangunan (PBB) merupakan salah satu sumber pendapatan asli daerah (PAD) yang memiliki peran strategis dalam pembiayaan pembangunan di tingkat desa[1]. Optimalisasi penerimaan PBB sendiri memerlukan strategi pemungutan yang efektif dan efisien[2]. Desa Gunasari, sebagai wilayah yang terus berkembang, menghadapi tantangan dalam mengelola data objek pajak yang tersebar secara geografis. Hal ini disebabkan karena nilai tanah dan bangunan sangat dipengaruhi oleh faktor aksesibilitas dan lokasi strategis, yang tidak selalu selaras dengan batas-batas administratif[3].

Faktor jalan merupakan indikator kunci yang mempengaruhi nilai properti dengan Jalan berperan sebagai infrastruktur utama yang menghubungkan suatu lokasi dengan pusat ekonomi, fasilitas umum dan akses transportasi[4]. Oleh karena itu, lokasi yang berdekatan dengan jalan utama atau jalan yang ber kondisi baik cenderung memiliki nilai jual dan daya tarik ekonomi lebih tinggi dibandingkan lokasi yang terisolasi [5]. Hal ini berimplikasi langsung pada nilai jual objek pajak (NJOP) dan potensi penerimaan PBB.

Penelitian berangkat dari asumsi bahwa pengelompokan (clustering) objek pajak PBB berdasarkan atribut jalan dapat memberikan gambaran yang lebih akurat mengenai potensi dan prioritas pemungutan PBB [6]. Analisis clustering memungkinkan pemerintah desa untuk mengidentifikasi zona-zona prioritas berdasarkan kesamaan karakteristik jalan dengan tujuan agar upaya intensifikasi dan ekstensifikasi pajak dapat lebih terarah, efektif, dan berbasis data[7]. Dengan demikian, strategi penagihan, sosialisasi, dan monitoring dapat dialokasikan secara proporsional, mulai dari wilayah dengan potensi tertinggi hingga yang memerlukan pendekatan khusus.

TINJAUAN LITERATUR

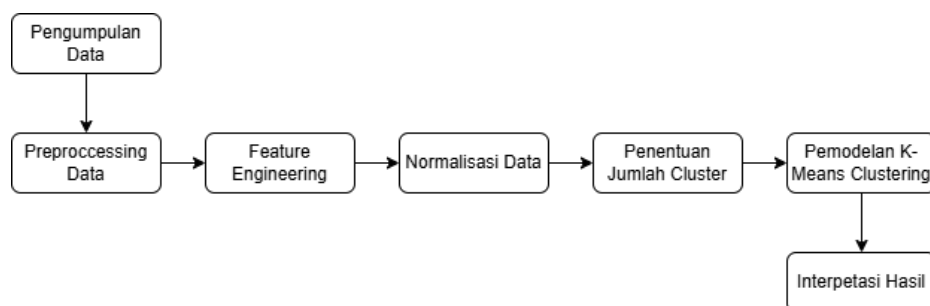
Pajak Keterkaitan antara anggaran pusat dan daerah memiliki peran penting dalam memperkuat kemandirian otonomi daerah[8]. Namun demikian, kapasitas fiskal daerah yang relatif kecil dibandingkan pemerintah pusat masih menjadi persoalan utama dalam pelaksanaan otonomi tersebut[9]. Menurut penelitian terbaru, desentralisasi fiskal yang tidak disertai dengan penguatan kapasitas keuangan daerah dapat memperburuk ketimpangan ekonomi antarwilayah dan menghambat efektivitas pemerintahan daerah. Upaya mencapai keseimbangan fiskal dilakukan melalui peningkatan pendapatan daerah, salah satunya melalui optimalisasi sumber-sumber keuangan daerah[10]. Berdasarkan fenomena tersebut, penelitian ini penting dilakukan untuk menganalisis faktor-faktor yang memengaruhi kepatuhan wajib pajak dalam membayar PBB-P2, khususnya pada daerah yang belum optimal dalam pencapaiannya[11].

Segmentasi yang digunakan dalam analisis PBB bertujuan untuk mengelompokkan wajib pajak ke dalam kelompok-kelompok tertentu agar strategi pengelolaan dan penanganan tunggakan dapat dilakukan secara lebih efektif[12]. Pada konteks PBB, segmentasi umumnya dilakukan berdasarkan wilayah, seperti kecamatan, desa, atau blok jalan, sebagaimana terlihat pada data jalan BL Babakan, BL Baginda, BL Bagogog, dan sebagainya. Pengelompokan berdasarkan wilayah ini membantu pemerintah atau instansi terkait dalam mengidentifikasi area yang memiliki tingkat tunggakan tinggi, wilayah yang patuh, serta area yang memerlukan perhatian khusus.

Dalam penelitian ini, metode K-means clustering digunakan sebagai solusi untuk mengelompokkan karakteristik objek PBB agar proses pengolahan data dalam jumlah besar menjadi lebih ringkas dan mudah dianalisis[13]. K-means dipilih karena mampu mengelompokkan objek-objek dengan kemiripan tertentu ke dalam cluster yang sama, sehingga memudahkan dalam mendeskripsikan pola dan karakteristik setiap kelompok wilayah atau wajib pajak[14]. Penerapan teknik data mining menggunakan teknik clustering telah terbukti efektif dalam meningkatkan efisiensi pengelolaan pajak daerah dengan memberikan segmentasi wajib pajak yang lebih akurat berdasarkan karakteristik geografis dan ekonomi[15]. Studi terbaru menunjukkan bahwa analisis clustering pada nilai jual objek pajak dapat membantu pemerintah daerah dalam mengidentifikasi potensi penerimaan yang belum teroptimalkan serta menetapkan strategi pemungutan yang lebih tepat sasaran[16]. Selain itu, pemanfaatan teknologi informasi dalam monitoring dan evaluasi penerimaan PBB telah menjadi kebutuhan mendesak di era digital dengan tujuan mengimplementasikan business intelligence dengan visualisasi data pajak secara real-time, sehingga pengambilan keputusan dapat dilakukan lebih cepat dan berbasis bukti empiris[17]. Hal ini sejalan dengan upaya modernisasi administrasi perpajakan yang bertujuan untuk meningkatkan transparansi, akuntabilitas, dan efektivitas pemungutan pajak daerah[18].

METODE PENELITIAN

Penelitian menggunakan pendekatan kuantitatif dengan metode analisis eksploratif. Penelitian termasuk dalam kategori penelitian terapan (*applied research*) yang bertujuan untuk memecahkan masalah praktis, yaitu menentukan prioritas pemungutan Pajak Bumi dan Bangunan (PBB).



Gambar 1. Alur Metode Penelitian

1. Pengumpulan data

Tahap pertama dalam penelitian adalah pengumpulan data Pajak Bumi dan Bangunan (PBB) bersumber dari sistem administrasi perpajakan Desa Gunasari. Data yang dikumpulkan meliputi informasi objek pajak berupa alamat objek pajak (nama jalan), tahun pajak, serta nilai PBB yang harus dibayarkan. Pada tahap ini, data yang diperoleh telah melalui proses anonimisasi sehingga tidak lagi memuat informasi identitas pribadi wajib pajak seperti Nomor Objek Pajak (NOP) dan nama wajib

pajak. Setelah data diperoleh, dilakukan proses verifikasi untuk memastikan kelengkapan dan keakuratan data, sehingga seluruh atribut yang dibutuhkan untuk analisis tersedia dan berada dalam kondisi valid. Selanjutnya, data didokumentasikan dan disimpan dalam format terstruktur guna memudahkan proses pengolahan, analisis, dan penerapan metode clustering pada tahap berikutnya. Data PBB yang digunakan dalam penelitian ini tergolong sebagai data sensitif, sehingga pengelolaannya dilakukan dengan memperhatikan prinsip etika dan tata kelola data. Seluruh data telah dianonimkan dan tidak memuat informasi identitas wajib pajak, seperti Nomor Objek Pajak (NOP) maupun nama wajib pajak. Analisis dilakukan pada data agregat tingkat jalan, sehingga tidak memungkinkan terjadinya identifikasi kembali terhadap individu tertentu. Akses dan penggunaan data dalam penelitian ini telah memperoleh izin dan otorisasi resmi dari pihak Pemerintah Desa Gunasari dan digunakan semata-mata untuk kepentingan penelitian dan perencanaan optimalisasi penagihan PBB.

2. Preprocessing Data

Pada tahap pra-pemrosesan, dilakukan pembersihan data untuk menangani permasalahan kualitas data yang ada. Proses dimulai dengan identifikasi struktur data dan penghapusan baris header yang tidak valid jika ditemukan. Kemudian dilakukan pembersihan karakter khusus pada kolom-kolom yang memiliki karakter tertentu, dimana karakter tersebut dihapus dan data dikonversi menjadi tipe numerik agar dapat diolah secara matematis. Pembersihan juga dilakukan pada nama kolom untuk menghilangkan karakter-karakter tidak perlu seperti tanda kutip, titik koma, dan backslash. Data nama jalan distandarisasi formatnya untuk memastikan konsistensi penulisan. Proses ini penting untuk memastikan bahwa data yang akan dianalisis memiliki kualitas yang baik dan bebas dari error yang dapat mempengaruhi hasil clustering.

3. Feature Engineering

Tahap ini melakukan agregasi data dari level wajib pajak individual menjadi level wilayah berdasarkan nama jalan. Perlu dicatat bahwa unit analisis dalam penelitian ini adalah tagihan PBB (*Objek Pajak*) bukan individu wajib pajak. Satu wajib pajak dapat memiliki lebih dari satu objek pajak misalnya, rumah dan tanah terpisah sehingga variabel 'Jumlah_Tagihan_Tertunggak' yang belum bayar setiap jalan, terlepas dari apakah beberapa tagihan tersebut dimiliki oleh wajib pajak yang sama. Pendekatan ini dipilih karena operasional penagihan dilakukan per objek pajak bukan per individu.

Setiap jalan dianalisis untuk menghasilkan metrik-metrik yang akan digunakan dalam clustering. Fitur pertama yang dibuat adalah jumlah tagihan tertunggak yang menghitung berapa banyak objek pajak yang belum dibayar PBB-nya pada setiap jalan, Fitur ini menggambarkan tingkat ketidakpatuhan pembayaran di wilayah tersebut. Fitur kedua adalah total PBB nilai PBB tertunggak yang merupakan akumulasi seluruh nilai pajak yang belum dibayar pada jalan tersebut, menunjukkan potensi penerimaan jika penagihan berhasil dilakukan.

Selain kedua fitur utama tersebut, dihitung juga statistik pendukung seperti rata-rata nilai PBB per tagihan, nilai tunggakan maksimum, dan nilai tunggakan minimum untuk memberikan gambaran distribusi nilai pajak di setiap wilayah. Meskipun fitur pendukung ini dihitung untuk keperluan analisis deskriptif dan interpretasi cluster, Pemodelan K-Means hanya menggunakan 2 Fitur utama jumlah tagihan tertunggak dan total nilai PBB tertunggak. Keputusan ini diambil karena kedua fitur tersebut secara langsung mencerminkan dua dimensi prioritas operasional (volume vs nilai), Fitur pendukung bersifat turunan dan berkorelasi tinggi dengan fitur utama dan prinsip parsimony untuk menjaga interpretabilitas model. Hasil agregasi ini menghasilkan dataset baru dimana setiap baris merepresentasikan satu jalan dengan karakteristik tunggakannya.

4. Normalisasi Data

Sebelum dilakukan clustering, kedua fitur utama yaitu jumlah tagihan tertunggak dan total nilai PBB tertunggak perlu dinormalisasi untuk menghindari bias yang disebabkan oleh perbedaan skala nilai. Normalisasi dilakukan menggunakan metode Z-Score Standardization (StandardScaler) yang mentransformasi data menjadi distribusi dengan rata-rata nol dan standar deviasi satu. Metode ini dipilih untuk menyetarakan skala kedua fitur dengan tujuan untuk memastikan bahwa jarak Euclidean dalam K-Means tidak didominasi oleh fitur dengan nilai numerik yang lebih besar.

5. Penentuan Jumlah Cluster Optimal

Pada penentuan jumlah cluster menggunakan dua metode validasi utama yaitu Elbow Method dan Silhouette Score. Elbow Method bekerja dengan menghitung inertia atau Within-Cluster Sum of Squares (WCSS) untuk berbagai nilai k dari 2 hingga 7, kemudian divisualisasikan dalam grafik untuk mencari titik "elbow" dimana penurunan inertia mulai melambat secara signifikan. Silhouette Score digunakan sebagai metode validasi kedua yang mengukur seberapa baik setiap objek ditempatkan dalam cluster-nya dibandingkan dengan cluster lain, dengan nilai berkisar antara -1 hingga 1 dimana nilai yang mendekati 1 menunjukkan clustering yang sangat baik. Kedua metode ini divisualisasikan secara berdampingan untuk memudahkan perbandingan dan pengambilan keputusan. Jumlah cluster optimal dipilih berdasarkan nilai Silhouette Score tertinggi, dengan mempertimbangkan juga interpretabilitas praktis untuk keperluan penentuan prioritas penagihan. Hasil evaluasi ditampilkan dalam tabel yang menunjukkan nilai inertia dan Silhouette Score untuk setiap k , beserta kategori kualitas

6. Implementasi Algoritma K-Means Clustering

Implementasi algoritma K-Means Clustering dilakukan menggunakan library scikit-learn versi 1.6.1 dengan python 3.12.12. Kemudian Algoritma K-Means Clustering ini dipilih karena sederhana, efisien untuk dataset berukuran menengah dan menghasilkan cluster yang mudah diinterpretasi untuk keperluan praktis penentuan prioritas operasional. Proses dimulai dengan inisialisasi awal centroid yang lebih optimal dibandingkan inisialisasi random.

Parameter yang digunakan meliputi: Algoritma='lloyd'; Init='k-means++'; random_state=42; n_init=10; max_iter=300. Kemudian pada setiap jalan di-assign ke cluster berdasarkan jarak euclidean dari fitur yang telah dinormalisasi dan label cluster ditambahkan ke dataset untuk analisis lebih lanjut.

7. Interpretasi Hasil Clustering

Visualisasi dibuat dalam empat panel yang saling melengkapi untuk memberikan gambaran komprehensif. Panel pertama adalah scatter plot yang menampilkan hubungan antara jumlah tunggakan dan total PBB tertunggak dengan pewarnaan berbeda untuk setiap kategori prioritas, memungkinkan identifikasi visual pola pengelompokan. Panel kedua adalah bar chart yang menunjukkan distribusi jumlah jalan pada setiap kategori prioritas untuk melihat proporsi wilayah di masing-masing tingkat prioritas. Panel ketiga adalah horizontal bar chart yang menampilkan 15 jalan dengan tunggakan tertinggi beserta kategori prioritasnya, memberikan informasi detail tentang wilayah-wilayah kritis yang memerlukan perhatian khusus. Panel keempat adalah pie chart yang memvisualisasikan proporsi kategori prioritas dalam bentuk persentase untuk memudahkan pemahaman distribusi secara keseluruhan. Setelah visualisasi, dilakukan analisis kuantitatif untuk setiap kategori prioritas yang meliputi jumlah jalan, total tagihan pbb tertunggak yang menunggak, total nilai PBB tertunggak, rata-rata tunggakan per jalan, dan tunggakan tertinggi. Analisis ini juga menyertakan contoh konkret jalan-jalan yang masuk dalam setiap kategori beserta nilai tunggakannya, sehingga memberikan insight praktis yang dapat langsung digunakan untuk perencanaan operasional penagihan pajak di Desa Gunasari.

HASIL PENELITIAN

Data PBB yang diperoleh dari sistem administrasi perpajakan Desa Gunasari terdiri dari atribut Nama jalan objek pajak, tahun pajak, serta nilai PBB terutang. Setelah proses verifikasi, Dataset mencakup 1383 tagihan PBB tertunggak yang tersebar di 106 jalan di Desa Gunasari seluruh data yang digunakan dipastikan valid dan lengkap sehingga dapat dilanjutkan ke tahap pra-pemrosesan. Pada tahap preprocessing, dilakukan standarisasi nama jalan untuk mengatasi inkonsistensi penulisan yang ditemukan dalam data. Sebelum standarisasi, dataset memiliki 105 nama jalan unik yang mencakup variasi penulisan tidak konsisten. Proses identifikasi menemukan 18 variasi penulisan yang mempengaruhi 56 tagihan, dengan pola inkonsistensi sebagai berikut: (1) penggunaan titik setelah prefix jalan sebanyak 15 kasus (contoh: 'BL. CIAWI' distandarisasi menjadi 'BL CIAWI'), (2) perbedaan format prefix sebanyak 2 kasus (contoh: 'BLOK CIAWI DARAT' distandarisasi menjadi 'BL CIAWI DARAT'), dan (3) penulisan kata tanpa spasi sebanyak 1 kasus (contoh: 'BL CIAWIDARAT' distandarisasi menjadi 'BL CIAWI DARAT'). Sebagai ilustrasi, jalan 'BL CIAWI DARAT' awalnya memiliki 4 variasi penulisan yang setelah distandarisasi menjadi satu format konsisten dengan total 57 tagihan. Proses standarisasi ini mengurangi jumlah jalan dari 105 menjadi 87

jalan unik, dimana pengurangan 18 jalan tersebut merupakan hasil penggabungan variasi penulisan inkonsisten yang merujuk pada jalan yang sama.

Tahap pra-pemrosesan dan pembentukan fitur menghasilkan dataset yang telah bersih, terstruktur, dan siap dianalisis pada tingkat wilayah. Pada tahap pra-pemrosesan, baris header ganda serta data tidak valid dihapus, karakter khusus seperti tanda titik koma (";") pada kolom nilai PBB dibersihkan, dan seluruh nilai PBB diubah menjadi tipe numerik agar dapat diproses secara matematis. Nama kolom dirapikan dengan menghapus karakter yang tidak diperlukan, sementara penulisan nama jalan distandarisasi untuk mencegah perbedaan format yang dapat mengganggu agregasi data. Dataset yang telah dibersihkan ini kemudian digunakan pada tahap feature engineering, di mana data objek pajak digabungkan berdasarkan nama jalan sehingga menghasilkan metrik-metrik agregat yang relevan. Perlu dicatat bahwa unit analisis dalam penelitian ini adalah tagihan PBB (objek pajak), bukan individu wajib pajak. Satu wajib pajak dapat memiliki lebih dari satu objek pajak (misalnya, rumah dan tanah terpisah), sehingga variabel yang dihitung adalah total tagihan PBB yang belum dibayar pada setiap jalan, terlepas dari apakah beberapa tagihan tersebut dimiliki oleh wajib pajak yang sama. Pendekatan ini dipilih karena operasional penagihan dilakukan per objek pajak, bukan per individu.

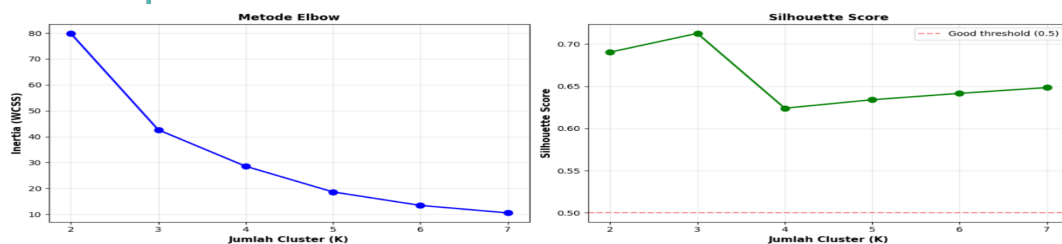
Fitur utama yang dibentuk meliputi jumlah tagihan PBB tertunggak per jalan dan total nilai rupiah tertunggak per jalan, yang menjadi indikator utama tingkat ketidakpatuhan dan potensi penerimaan. Selain itu dihitung pula fitur pendukung seperti rata-rata nilai tunggakan, tunggakan tertinggi, dan tunggakan terendah pada masing-masing jalan untuk memberikan gambaran lebih mendalam mengenai karakteristik distribusi tunggakan di setiap wilayah. Hasil akhir kedua tahap ini ditunjukkan pada Tabel 1 (10 baris pertama), yang memperlihatkan bahwa data telah berada dalam kondisi bersih dan telah teragregasi dengan baik sehingga siap digunakan untuk proses normalisasi dan analisis clustering pada tahap berikutnya.

Tabel 1. Data PBB Setelah Praprocess

Jalan	Jumlah_Tunggakan	Total_PBB_Tertunggak
BL ASEM	37	7927429.0
BL CITALANG	22	7254279.0
BL CIPONGPORANG	13	6769764.0
BL GANEAS	19	6173185.0
BL SUKAMULYA	36	3771533.0
BL CIMANDALA	43	3741235.0
BL BOGOGOG	59	3440239.0
BL GUNUNGGAJAH	50	3431088.0
BL CIHONJE	41	3417570.0
BL PANGANGONAN	2	3246096.0

Pada tahap normalisasi data, dua fitur utama yang digunakan dalam analisis clustering, yaitu Jumlah_Tagihan_Tertunggak dan Total_Nilai_Tertunggak, terlebih dahulu dipisahkan dan dibersihkan dari nilai kosong dengan menggantinya menggunakan nilai nol. Kedua fitur tersebut kemudian dinormalisasi menggunakan metode StandardScaler sehingga setiap variabel memiliki distribusi dengan rata-rata nol dan standar deviasi satu. Proses normalisasi ini dilakukan untuk menghilangkan perbedaan skala antar fitur yang berpotensi menimbulkan bias pada perhitungan jarak Euclidean dalam algoritma K-Means.

Dataset yang telah distandarisasi selanjutnya digunakan dalam proses penentuan jumlah cluster optimal. Evaluasi dilakukan dengan menghitung nilai inertia (Within-Cluster Sum of Squares) dan Silhouette Score pada rentang $k = 2$ hingga $k = 7$. Nilai inertia menunjukkan pola penurunan seiring bertambahnya jumlah cluster, dan divisualisasikan melalui grafik Elbow Method untuk mencari titik belokan (elbow) sebagai indikasi jumlah cluster yang efisien. Di sisi lain, Silhouette Score dihitung untuk menilai kualitas pemisahan cluster, dimana nilai yang lebih tinggi menunjukkan struktur pengelompokan yang lebih baik. Kedua hasil evaluasi tersebut divisualisasikan secara berdampingan sehingga memudahkan identifikasi jumlah cluster yang paling sesuai untuk digunakan dalam analisis. Nilai k optimal kemudian dipilih berdasarkan kombinasi penurunan inertia yang stabil dan nilai Silhouette Score tertinggi yang diperoleh.



Gambar 2. Metode Elbow dan Silhouette Score

Tabel 2 Evaluasi Jumlah Cluster

K	Inertia	Silhouette	Kualitas
2	67.44	0.648	Good
3	34.28	0.681	Good
4	22.45	0.616	Good
5	14.90	0.597	Good
6	10.93	0.614	Good
7	8.59	0.618	Good

Evaluasi jumlah cluster melalui Elbow Method dan Silhouette Score menunjukkan bahwa meskipun inertia terus menurun seiring bertambahnya jumlah cluster, titik belokan grafik berada pada kisaran $k = 3-4$. Nilai Silhouette Score tertinggi terdapat pada $k = 3$ dengan skor 0,681 yang dikategorikan Good, lebih tinggi dibandingkan nilai pada $k = 2$ (0,648), $k = 4$ (0,616), dan k lainnya yang berada pada kategori Good dengan nilai lebih rendah.

Pemilihan $k = 3$ dibandingkan $k = 4$ didasarkan pada beberapa pertimbangan: (1) perbedaan Silhouette Score yang signifikan (0,681 vs 0,616), mengindikasikan kualitas pemisahan cluster yang lebih baik pada $k = 3$; (2) penurunan inertia yang sudah melambat setelah $k = 3$ (dari 34,28 ke 22,45), menunjukkan tambahan cluster tidak memberikan perbaikan substansial; (3) interpretabilitas praktis untuk kebijakan penagihan, dimana tiga kategori prioritas (rendah-menengah-tinggi) lebih mudah diimplementasikan secara operasional dibandingkan empat kategori yang berpotensi membingungkan. Dengan demikian, kombinasi penurunan inertia yang signifikan dan kualitas pemisahan cluster terbaik mendukung bahwa $k = 3$ merupakan jumlah cluster optimal untuk digunakan dalam pengelompokan prioritas penagihan PBB di Desa Gunasari.

Berdasarkan hasil implementasi K-Means dengan $k = 3$, algoritma berhasil konvergen setelah 3 iterasi, menghasilkan tiga kelompok prioritas penagihan PBB dengan karakteristik yang berbeda. Jalan-jalan di Desa Gunasari terbagi menjadi: Cluster 0 (Prioritas Rendah) dengan 62 jalan (71,3%), Cluster 1 (Prioritas Menengah) dengan 21 jalan (24,1%), dan Cluster 2 (Prioritas Tinggi) dengan 4 jalan (4,6%). Pembagian ini membantu menentukan fokus operasional penagihan, mulai dari wilayah yang membutuhkan perhatian intensif hingga yang tergolong patuh. Rincian masing-masing cluster dapat dilihat pada Tabel 3.

Tabel 3. Implementasi K-Means

Cluster	Kategori	Jumlah Jalan	Rata Rata Tunggakan WP	Rata-Rata total PBB
0	Prioritas Rendah	62 jalan	6,4 Orang	Rp 458,544
1	Prioritas Menengah	21 jalan	42,2 orang	Rp 2,299,109
2	Prioritas Tinggi	4 jalan	25,2 orang	Rp 7,328.102

Berdasarkan tabel diatas, strategi prioritas penagihan PBB dapat difokuskan sesuai karakteristik masing-masing cluster:

Cluster 0 (Prioritas Rendah) mencakup sebagian besar jalan yaitu 62 jalan dengan rata-rata 6,4 tagihan tertunggak per jalan dan rata-rata nilai tertunggak sebesar Rp 458.544. Wilayah ini menunjukkan tingkat kepatuhan yang tinggi, dimana sebagian besar objek pajak terbayar secara rutin, sehingga fokus penagihan dapat bersifat preventif dan periodik tanpa memerlukan intervensi khusus.

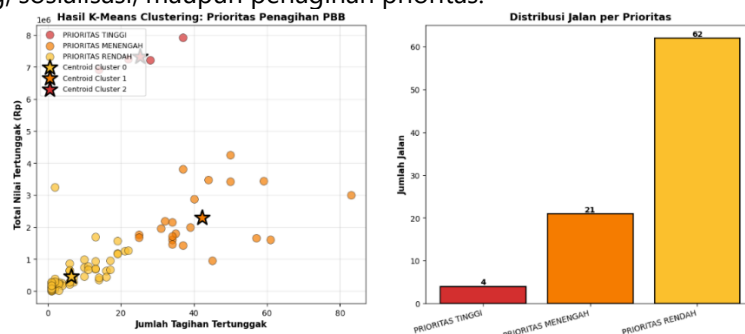
Cluster 1 (Prioritas Menengah) terdiri dari 21 jalan dengan rata-rata 42,2 tagihan tertunggak per jalan, jumlah tertinggi diantara ketiga cluster dan rata-rata nilai tertunggak sebesar Rp 2.299.109. Jumlah tagihan tertunggak yang sangat tinggi (42,2 vs 6,4 di cluster rendah) menunjukkan pola ketidakpatuhan massal yang memerlukan pengawasan lebih intensif, penjadwalan penagihan yang terstruktur, dan strategi sosialisasi untuk memastikan kepatuhan.

Cluster 2 (Prioritas Tinggi) walaupun jumlah jalannya sedikit yakni 4 jalan dengan rata-rata 25,2 tagihan tertunggak, memiliki nilai rata-rata tertunggak yang sangat besar yaitu Rp 7.328.102 per jalan—hampir 16 kali lipat nilai cluster rendah dan 3 kali lipat nilai cluster menengah. Nilai tertunggak yang sangat tinggi ini mengindikasikan adanya objek pajak dengan NJOP tinggi (properti premium atau komersial) yang menunggak, menjadikannya fokus utama intervensi penagihan karena kontribusi finansialnya yang sangat signifikan terhadap potensi penerimaan PBB desa. Pembagian prioritas ini memungkinkan pengelolaan sumber daya penagihan PBB lebih efisien dan efektif, serta membantu pemerintah desa dalam menetapkan strategi intervensi berdasarkan risiko finansial masing-masing wilayah. pemerintah desa dalam menetapkan strategi intervensi berdasarkan risiko finansial masing-masing wilayah.

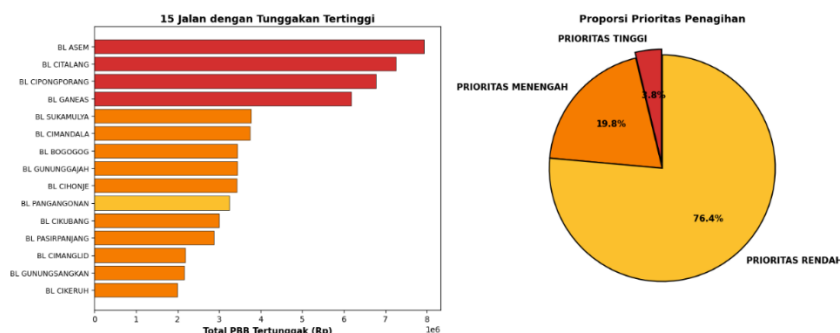
Hasil clustering dengan $k=3$ memberikan pemahaman yang jelas mengenai pola tunggakan PBB di Desa Gunasari. Cluster 0 (Prioritas Rendah) menunjukkan wilayah yang relatif patuh, dengan jumlah jalan terbanyak (71,3%) namun tingkat tunggakan dan total PBB tertunggak rendah. Hal ini menandakan bahwa sebagian besar objek pajak di cluster ini terbayar secara rutin, sehingga fokus penagihan di wilayah ini dapat bersifat preventif dan periodik.

Cluster 1 (Prioritas Menengah) mencerminkan wilayah dengan kepatuhan menengah, ditandai oleh jumlah tagihan tertunggak yang sangat tinggi meskipun total nilai PBB tertunggak tidak setinggi cluster prioritas tinggi. Wilayah ini memerlukan perhatian lebih melalui penjadwalan penagihan yang terstruktur dan monitoring kepatuhan yang lebih intensif.

Sementara itu, Cluster 2 (Prioritas Tinggi) walaupun jumlah jalannya sedikit (hanya 4,6% dari total jalan), memiliki total nilai PBB tertunggak yang sangat besar, sehingga memberikan kontribusi signifikan terhadap potensi penerimaan pajak. Wilayah ini menjadi fokus utama intervensi, baik melalui pendekatan komunikasi langsung, sosialisasi, maupun penagihan prioritas.



Gambar 3. Hasil K-Means & Distribusi Jalan



Gambar 4. Top 15 Jalan & Proporsi Prioritas

Hasil pengelompokan menunjukkan pola yang jelas antara jumlah tagihan tertunggak dan total nilai PBB tertunggak di setiap jalan. Sebagian besar jalan berada pada kategori prioritas rendah dengan tunggakan dan nilai PBB relatif kecil, sementara jalan-jalan dengan prioritas menengah memiliki jumlah tagihan tertunggak yang tinggi meskipun total nilai PBB-nya sedang. Kelompok prioritas tinggi meski terdiri dari sedikit jalan, memiliki nilai PBB tertunggak yang sangat besar, sehingga menjadi fokus utama penagihan. Pola ini memperlihatkan bahwa distribusi tunggakan tidak merata, sehingga strategi penagihan dapat diarahkan berdasarkan tingkat risiko dan potensi penerimaan tiap wilayah.

PEMBAHASAN

Hasil penelitian menunjukkan bahwa pengelompokan jalan di Desa Gunasari menggunakan K-Means dengan $k = 3$ berhasil mengidentifikasi pola tunggakan PBB secara jelas. Evaluasi model dengan kombinasi Elbow Method dan Silhouette Score mendukung pemilihan jumlah cluster optimal. Nilai Silhouette Score sebesar 0,681 untuk $k = 3$ mengindikasikan bahwa cluster yang terbentuk memiliki pemisahan yang baik dan homogenitas internal yang cukup tinggi, sehingga hasil clustering dapat dijadikan dasar yang andal untuk pengambilan keputusan operasional.

Meskipun nilai Silhouette Score berada pada kategori Good (bukan Excellent yang memerlukan nilai $>0,70$), hasil ini merupakan pencapaian yang wajar dan bahkan baik untuk data geografis dengan heterogenitas tinggi seperti data PBB tingkat desa. Data PBB memiliki karakteristik natural dimana objek pajak dalam satu jalan dapat memiliki nilai yang sangat beragam—dari rumah sederhana hingga properti komersial—sehingga overlap antar cluster merupakan karakteristik inherent dari data, bukan indikasi kualitas clustering yang buruk. Nilai ini konsisten dengan penelitian serupa pada domain perpajakan daerah [13, 16] yang umumnya menghasilkan Silhouette Score berkisar 0,62-0,70. Lebih penting lagi, hasil clustering menunjukkan separasi nilai yang sangat jelas secara praktis dengan rasio 1:5:16 (rendah:menengah:tinggi) dan distribusi yang masuk akal operasional (71%-24%-5%), mengindikasikan bahwa cluster yang terbentuk memiliki makna bisnis yang kuat.

Cluster 2 (Prioritas Tinggi) - Strategi Value-Based: Meski jumlah tagihan per jalan lebih rendah (25,2) dibanding Cluster 1 (42,2), memiliki nilai rata-rata tertunggak yang jauh lebih besar (Rp 7.328.102 vs Rp 2.299.109). Ini mengindikasikan adanya objek pajak dengan NJOP tinggi (properti premium atau komersial) yang menunggak, dimana setiap tagihan yang berhasil ditagih memberikan kontribusi finansial sangat signifikan. Strategi operasional: Penagihan prioritas dengan pendekatan personal dan komunikasi intensif. Tim penagihan dapat mengalokasikan petugas senior untuk melakukan koordinasi langsung dengan pemilik properti bernilai tinggi.

Cluster 1 (Prioritas Menengah) - Strategi Volume-Based: Memiliki jumlah tagihan tertunggak tertinggi (42,2 per jalan) namun nilai rata-rata lebih rendah, mencerminkan volume tunggakan massal pada properti dengan NJOP sedang. Meskipun nilai per tagihan tidak sebesar Cluster 2, akumulasi volume yang tinggi tetap memberikan potensi penerimaan substansial. Strategi operasional: Sosialisasi berkelanjutan, penjadwalan penagihan terstruktur secara berkala, dan program insentif pembayaran (misalnya keringanan denda) untuk mendorong kepatuhan massal. Pendekatan ini lebih efisien dibanding penagihan individual mengingat volume yang besar.

Cluster 0 (Prioritas Rendah) - Strategi Pemeliharaan: Volume dan nilai rendah (6,4 tagihan, Rp 458.544), menunjukkan kepatuhan baik di sebagian besar wilayah (71,3% jalan). Strategi operasional: Pemeliharaan rutin melalui reminder berkala (SMS/surat), layanan pembayaran yang mudah diakses, dan apresiasi kepada wajib pajak patuh untuk mempertahankan tingkat kepatuhan tinggi.

Dari evaluasi model, inertia yang menurun tajam dari $k = 2$ ke $k = 3$ (dari 67.44 menjadi 34.28) menunjukkan bahwa peningkatan cluster memberikan perbaikan signifikan dalam pengelompokan, sedangkan setelah $k > 3$ penurunan inertia melambat (dari 34.28 ke 22.45 untuk $k=4$), menandakan tambahan cluster tidak banyak meningkatkan efektivitas pemisahan. Kombinasi nilai inertia dan Silhouette Score memastikan bahwa $k = 3$ bukan hanya optimal secara matematis, tetapi juga praktis untuk strategi penagihan berbasis prioritas. Tiga kategori prioritas juga lebih mudah dikomunikasikan kepada tim lapangan dan stakeholder dibandingkan dengan pembagian yang lebih kompleks.

Implementasi clustering ini memberikan panduan strategis bagi pemerintah desa dalam alokasi sumber daya. Jalan-jalan di cluster prioritas tinggi harus menjadi target utama penagihan karena kontribusi tunggaknya besar terhadap total potensi penerimaan. Sementara itu, cluster prioritas menengah membutuhkan monitoring dan penjadwalan penagihan yang sistematis untuk mengatasi volume tunggakan yang tinggi, dan cluster prioritas rendah dapat ditangani secara rutin dengan intervensi minimal namun konsisten. Strategi berbasis cluster ini memungkinkan penagihan lebih efisien, mengurangi risiko kehilangan penerimaan, dan meningkatkan kepatuhan wajib pajak secara keseluruhan.

Selain itu, hasil visualisasi memperkuat interpretasi clustering, menunjukkan distribusi yang tidak merata dan mengidentifikasi jalan-jalan kritis dengan tunggakan tinggi. Scatter plot pada Gambar 3 memperlihatkan pemisahan yang jelas antar cluster, sementara analisis Top 15 jalan pada Gambar 4 memberikan informasi spesifik untuk tindakan cepat. Dengan demikian, pendekatan berbasis K-Means ini

tidak hanya valid secara statistik, tetapi juga relevan secara operasional, memungkinkan desa membuat keputusan berbasis data untuk pengelolaan PBB yang lebih efektif dan terukur.

KESIMPULAN

Penggunaan metode K-Means dengan pendekatan kuantitatif efektif dalam mengelompokkan jalan di Desa Gunasari berdasarkan jumlah tagihan PBB tertunggak dan total nilai tertunggak. Penelitian ini berhasil menganalisis 1.382 tagihan PBB tertunggak yang tersebar di 87 jalan untuk periode pajak tahun 2024. Evaluasi model dengan Elbow Method dan Silhouette Score menunjukkan bahwa jumlah cluster optimal adalah $k = 3$, dengan nilai Silhouette Score 0,681 yang menunjukkan kualitas pemisahan cluster yang baik (kategori Good) dan homogenitas internal yang cukup tinggi. Hal ini memastikan bahwa pengelompokan dapat digunakan sebagai dasar pengambilan keputusan operasional yang andal.

Tiga cluster yang terbentuk mencerminkan tingkat prioritas penagihan yang berbeda dengan strategi operasional yang spesifik. Cluster 0 (Prioritas Rendah) terdiri dari sebagian besar jalan (62 jalan, 71,3%) dengan rata-rata 6,4 tagihan tertunggak dan nilai rata-rata Rp 458.544 per jalan, sehingga penagihan di wilayah ini dapat dilakukan secara rutin dan preventif. Cluster 1 (Prioritas Menengah) memiliki 21 jalan (24,1%) dengan rata-rata 42,2 tagihan tertunggak per jalan—jumlah tertinggi—dan nilai rata-rata Rp 2.299.109, memerlukan pengawasan lebih intensif dan penjadwalan penagihan yang sistematis dengan pendekatan sosialisasi massal (volume-based approach). Cluster 2 (Prioritas Tinggi), meski jumlah jalannya sedikit (4 jalan, 4,6%) dengan rata-rata 25,2 tagihan tertunggak, memiliki nilai rata-rata tertunggak yang sangat besar yaitu Rp 7.328.102 per jalan, sehingga menjadi fokus utama intervensi penagihan dengan pendekatan personal dan komunikasi intensif (value-based approach).

Hasil clustering menunjukkan distribusi tunggakan yang tidak merata, mengindikasikan bahwa strategi penagihan berbasis prioritas lebih efektif dibandingkan strategi seragam. Pendekatan berbasis cluster memungkinkan alokasi sumber daya yang lebih efisien melalui diferensiasi strategi: value-based approach untuk cluster prioritas tinggi yang fokus pada nilai nominal besar, volume-based approach untuk cluster prioritas menengah yang menangani ketidakpatuhan massal, dan maintenance approach untuk cluster prioritas rendah yang mempertahankan kepatuhan tinggi. Dengan demikian, metode K-Means terbukti tidak hanya valid secara statistik tetapi juga relevan secara praktis untuk mendukung pengambilan keputusan operasional dan perencanaan strategi penagihan PBB di Desa Gunasari.

REFERENCES

- [1] G. Angelita Tarigan, J. Miechell Liem, M. Samy Abdul Rahman Shebubakar, and M. Cahaya, "Kontribusi Pajak Bumi dan Bangunan (PBB) dan Bea Perolehan Hak Atas Tanah dan Bangunan (BPHTB) Terhadap Pendapatan Asli Daerah (PAD) di Kota Tangerang," *J. Account. Financ. Manag.*, vol. 6, no. 2, pp. 550–566, 2025, doi: 10.38035/jafm.v6i2.1856.
- [2] Audi Fazhillah and Nurlaila Nurlaila, "Analisis Penerimaan Pajak Bumi dan Bangunan (PBB) terhadap Pendapatan Daerah di Kecamatan Kotarih," *J. Ris. Akunt.*, vol. 2, no. 2, pp. 22–35, 2024, doi: 10.54066/jura-itb.v2i2.1694.
- [3] N. Sebagai, P. Pbb, D. I. Kota, A. Syaharani, I. Ismatullah, and R. Nurmilah, "ANALISIS FAKTOR – FAKTOR YANG PENENTUAN NILAI JUAL OBJEK PAJAK (Studi Kasus pada Kecamatan Gunungpuyuh)," vol. 10, no. 204, pp. 2256–2267, 2025.
- [4] A. Purwitasari, B. Mutafarida, and Yuliani, "Urgensi Pajak dalam Mendorong Pembangunan Infrastruktur dan Pertumbuhan Ekonomi di Indonesia," *J. Ilm. Ekon. dan Manaj.*, vol. 2, no. 6, pp. 656–666, 2024, [Online]. Available: <https://doi.org/10.61722/jiem.v2i6.1584>
- [5] A. Z. M. H. Sulastris, "Journal of Law, Administration, and Social Science Volume 1 No. 1, Juni 2021 ANALISIS PENGARUH JALAN BYPASS MATARAM-GERUNG TERHADAP NILAI TANAH DI LOMBOK BARAT," vol. 1, no. 1, pp. 40–46, 2021.
- [6] R. Magriaty, K. Murtiaksiono, and S. Anwar, "Analisis K-Means Cluster untuk Identifikasi Kawasan Pengelolaan Sampah di Kabupaten Tapin Provinsi Kalimantan Selatan," *J. Reg. Rural Dev. Plan.*, vol. 7, no. 1, pp. 79–90, 2023, doi: 10.29244/jp2wd.2023.7.1.79-90.
- [7] N. N. Adisanjaya, A. T. A. Prawira Kusuma, and I. G. N. Manik Nugraha, "Pemetaan Zonasi Daerah Rawan Banjir di Denpasar Bali dengan Metode K-MEANS CLUSTERING," *J. Media Sains*, vol. 5, no. 1, pp. 37–46, 2021, doi: 10.36002/jms.v5i1.1495.

- [8] F. V. Valentina, V. Alfiyan, and M. I. Anshori, "Jurnal Ilmu Manajemen, Bisnis dan Ekonomi | JIMBE <https://malaqbipublisher.com/index.php/JIMBE>," *J. Ilmu Manajemen, Bisnis dan Ekon. JIMBE*, vol. 1, no. 5, pp. 285–292, 2024.
- [9] Arta Tio Pratama, Diya Gustiara Putri, Mila Mayasari, and Anas Malik, "Tantangan Dan Peluang Otonomi Daerah Dalam Mendukung Keberlanjutan Pertumbuhan Ekonomi Lokal," *Kalianda Halok Gagas*, vol. 8, no. 1, pp. 1–16, 2024, doi: 10.52655/khg.v8i1.103.
- [10] S. Mu'minah and M. B. Z. Tjenreng, "Desentralisasi dan Ketimpangan Pembangunan Antar Daerah," *Sci. J. Reflect. Econ. Accounting, Manag. Bus.*, vol. 8, no. 1, pp. 342–351, 2025, doi: 10.37481/sjr.v8i1.1053.
- [11] N. Wulandari and D. Wahyudi, "... Pengetahuan Perpajakan, Sanksi Pajak, Kesadaran Wajib Pajak, Dan Kualitas Pelayanan Pajak Terhadap Kepatuhan Wajib Pajak Dalam Membayar Pajak Bumi ..., " *J. ...*, vol. 6, pp. 14853–14870, 2022.
- [12] D. Kurniadi *et al.*, "(1) , Yoga.Handoko@Itg.Ac.Id 2) , 1806129@Itg.Ac.Id 3) , Idafarida@Itg.Ac," *AITI J. Teknol. Inf.*, vol. 20, no. 1, pp. 64–77, 2023.
- [13] V. No, A. M. Nur, H. Bahtiar, M. A. Jannah, and V. No, "Infotek: Jurnal Informatika dan Teknologi Implementasi Algoritma K-Means Clustering Dalam Mengelompokkan Kepatuhan Wajib Pajak Bumi dan Bangunan Dengan Optimasi Elbow Sebuah negara dikatakan maju jika mampu membangun wilayah dan masyarakatnya , baik seca," *Infotek J. Inform. dan Teknol.*, vol. 8, no. 1, 2025.
- [14] A. A. Alya Putri and S. A. Rahmah, "Implementasi Data Mining Dengan Algoritma K-Means Clustering Untuk Analisis Bisnis Pada Perusahaan Asuransi," *Djtechno J. Teknol. Inf.*, vol. 5, no. 1, pp. 139–152, 2024, doi: 10.46576/djtechno.v5i1.4537.
- [15] F. Corly and S. Ariana, "Peran Big Data Dan Artificial Intelligence," *Peringkat Akreditasi Sinta*, vol. 23, no. 3, pp. 2655–826, 2024.
- [16] R. Cantika, A. Fauzi, and A. Sihombing, "Penerapan Metode Clustering Untuk Mengetahui Kepatuhan Wajib Pajak Bumi Dan Bangunan Pada Desa Perkebunan Tanjung Keliling kemampuannya dalam memajukan kesejahteraan masyarakat . Untuk mencapai tujuan yang dan / atau perairan pedalaman dan / atau laut . A," *J. Publ. Sist. Inf. dan Telekomun.*, vol. 2, no. 4, 2024.
- [17] S. Pranoto and S. Wahyuni, "Integrasi Data Mining Dan Business Intelligence Monitoring Piutang Pbb-P2 Bpkpd Kota Tebing Tinggi," *J. Sci. Soc. Res.*, vol. 4307, no. 2, pp. 1213–1222, 2025, [Online]. Available: <http://jurnal.goretanpena.com/index.php/JSSR>
- [18] K. Khalida, M. Riska, and N. A. Ajianing, "Strategi Efektivitas Pengumutan Pajak Daerah Dan Pajak Negera Untuk Pembangunan Ekonomi Berkelanjutan," *J. Media Akad.*, vol. 3, no. 6, pp. 3031–5220, 2025, [Online]. Available: <https://jurnal.mediaakademik.com/index.php/jma/article/view/1999>