

Identifikasi Tingkat Kepatuhan Wajib Pajak Bumi dan Bangunan di Jakarta dengan Pendekatan Algoritma *K-Means*

Muhammad Taufiqul Umam^{1*}, Baby Lolita Basyah²

^{1,2}Magister Manajemen Sistem Informasi, Universitas Gunadarma, Jl. Margonda Raya No. 100, Kota Depok, Jawa Barat, 16424, Indonesia

E-mail: ¹taufiqulmam@gmail.com, ²b_lolita@staff.gunadarma.ac.id

*Corresponding Author

Abstrak— PBB-P2 merupakan instrumen fiskal daerah yang dikenakan atas hak kepemilikan, wewenang penguasaan, maupun kemanfaatan ekonomi dari bumi dan bangunan oleh subjek pajak, baik perseorangan maupun entitas hukum. Sebagai salah satu pilar utama perpajakan daerah, PBB-P2 memegang peranan strategis dalam memperkuat struktur Pendapatan Asli Daerah (PAD). Jakarta, dengan posisinya sebagai episentrum ekonomi dan pemerintahan yang memiliki densitas penduduk serta aktivitas bisnis yang masif, menyimpan potensi penerimaan PBB-P2 yang sangat strategis untuk dikelola secara maksimal. Penelitian ini berfokus pada pemetaan derajat kepatuhan wajib pajak di Jakarta melalui penerapan algoritma *K-Means*. Prosedur penelitian ini mengikuti kerangka kerja *Cross-Industry Standard Process for Data Mining (CRISP-DM)* yang terdiri dari fase *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, serta *deployment*. Analisis dilakukan terhadap 28.125 data wajib pajak PBB-P2 selama kurun waktu 2020 hingga 2024. Temuan studi menunjukkan terbentuknya empat kategori kelompok, di mana *cluster 0* merepresentasikan tingkat kepatuhan sangat tinggi, *cluster 1* untuk tingkat rendah, *cluster 2* untuk kategori sedang, dan *cluster 3* mencerminkan tingkat kepatuhan yang tinggi. Dengan perolehan skor *silhouette coefficient* mencapai 0.742, hasil segmentasi ini terbukti memiliki kualitas yang valid, yang mengindikasikan bahwa setiap anggota dalam kelompok memiliki karakteristik serupa dan terlihat kontras dibandingkan kelompok lainnya.

Kata Kunci — Algoritma *K-Means*; Clustering; Kepatuhan Wajib Pajak; Pajak Bumi dan Bangunan.

Abstract— The Rural and Urban Land and Building Tax (PBB-P2) serves as a regional fiscal mechanism levied on the ownership, control, or utilization of property assets by both individuals and corporate entities. As a cornerstone of regional fiscal policy, PBB-P2 is instrumental in bolstering Local Own-Source Revenue (PAD). Given Jakarta's status as the nation's administrative and commercial epicenter characterized by high population density and intense economic momentum, the city holds a strategic and vast potential for PBB-P2 collection. This study aims to categorize the compliance behavior of PBB-P2 taxpayers within the Jakarta region by utilizing the *K-Means* algorithm. The research methodology is guided by the *Cross-Industry Standard Process for Data Mining (CRISP-DM)* framework, which involves six systematic phases: *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, and *deployment*. The dataset consists of 28,125 PBB-P2 taxpayer records collected from 2020 to 2024. The findings reveal that taxpayer compliance is classified into four distinct clusters. Cluster 0 indicates very high compliance, Cluster 1 reflects low compliance, Cluster 2 denotes moderate compliance, and Cluster 3 corresponds to high compliance. An average silhouette coefficient value of 0.742 demonstrates that the resulting clusters are well-defined, showing strong internal similarity and clear separation from one another.

Keywords — *K-Means* Algorithm; Clustering; Taxpayer Compliance; Land and building Tax.

I. PENDAHULUAN

Selaras dengan amanat regulasi Pemerintahan Nomor 35 2023 dalam hal aturan Umum Pajak Daerah serta Retribusi Daerah, Pajak Daerah dimaksudkan sebagai kontribusi wajib yang dibebankan kepada wajib pajak orang pribadi maupun badan usaha. Kontribusi ini dikelola oleh pemerintah daerah melalui mekanisme pemungutan yang bersifat memaksa (*koersif*) dengan berpijak pada legitimasi hukum yang berlaku. Pelaksanaan pemungutan pajak ini dilakukan tanpa memberikan imbal balik atau imbalan langsung kepada pembayar pajak, melainkan difokuskan sepenuhnya untuk membiayai belanja kepentingan umum serta mengakselerasi peningkatan taraf kesejahteraan masyarakat secara luas. Dalam struktur kehidupan bernegara, instrumen pajak menempati posisi yang sangat krusial, khususnya sebagai pilar utama pembiayaan dalam penyelenggaraan pemerintahan serta pengadaan berbagai kebutuhan strategis negara [1]. Pajak Bumi serta Bangunan Perdesaan dan juga Perkotaan (PBB-P2) merupakan instrumen fiskal daerah yang memegang peranan krusial dalam mengeskalasi Pendapatan Asli Daerah

(PAD) secara optimal. Beban pajak ini ditetapkan atas objek pajak berupa lahan dan konstruksi bangunan yang berada di bawah kepemilikan, kendali, maupun pemanfaatan oleh subjek pajak terkait [2].

Sebagai wilayah yang menjadi sentra aktivitas pemerintahan dan ekonomi nasional, Provinsi DKI Jakarta memiliki prospek yang sangat signifikan dalam pelaksanaan pengadopsian Pajak Bumi serta Bangunan Perdesaan dan juga Perkotaan (PBB-P2). Meskipun demikian, capaian penerimaan PBB-P2 selama periode 2020 hingga 2024 dalam pelaksanaannya masih menunjukkan dinamika yang berfluktuasi secara cukup signifikan. Data dari Badan Pendapatan Daerah DKI Jakarta memperlihatkan bahwa hanya pada tahun 2023 target realisasi tercapai, sedangkan tahun-tahun lainnya masih berada di bawah target. Situasi ini menunjukkan adanya hambatan terkait kedisiplinan wajib pajak dalam memenuhi tanggung jawab pembayaran PBB-P2, baik ditinjau dari sisi akurasi waktu penyetoran maupun kepatuhan terhadap regulasi yang berlaku. Menanggapi permasalahan tersebut, perumusan strategi yang lebih akurat dan efisien sangat mendesak untuk dilakukan guna memacu peningkatan realisasi setoran PBB-P2, sehingga potensi Pendapatan Asli Daerah (PAD) dapat digali dan dimanfaatkan secara maksimal. Upaya ini menjadi krusial guna memastikan kemampuan pemerintah daerah dalam mengelola serta mengoptimalkan sumber-sumber pajak daerah secara berkelanjutan [3].

Kepatuhan perpajakan merefleksikan sikap serta kesadaran subjek pajak dalam menjalankan tanggung jawab sekaligus mengelola hak-hak fiskal mereka, selaras dengan parameter yang telah diatur dalam kerangka peraturan perundang-undangan [4]. Derajat kepatuhan para wajib pajak menjadi elemen fundamental dalam memacu pertumbuhan perolehan pajak wilayah. Terdapat beberapa variabel yang teridentifikasi memberikan pengaruh positif terhadap kedisiplinan penyetoran PBB-P2, di antaranya adalah kedalaman pemahaman subjek pajak mengenai regulasi fiskal, pemberlakuan sanksi bagi pelanggar aturan, serta efisiensi dalam proses distribusi Surat Pemberitahuan Pajak Terutang (SPPT) [5]. Namun, faktor-faktor tersebut belum cukup efektif menjelaskan seluruh variasi taat wajib pajak di Jakarta. Maka dibutuhkan pendekatan yang lebih analitis dan berbasis data untuk memahami pola-pola kepatuhan wajib pajak secara lebih komprehensif.

Implementasi teknik *data mining* menjadi salah satu solusi strategis dalam menanggapi persoalan tersebut. Pendekatan ini dioptimalkan untuk mengeksekusi proses pengolahan, kategorisasi, serta pemeriksaan mendalam terhadap basis data berskala besar guna mengekstraksi pola-pola tersirat yang memiliki relevansi kuat dan nilai informasi yang berharga [6]. Dalam konteks tersebut, teknik *clustering* dinilai sebagai metodologi yang relevan untuk mengklasifikasikan wajib pajak berdasarkan derajat dan tipologi kepatuhan mereka. Di antara berbagai metode yang ada, algoritma *K-Means* ditandai sebagai salah satu teknik pengelompokan yang paling sering diadopsi dan diakui efektivitasnya dalam berbagai kajian literatur. Mekanisme metode ini berfokus pada pengorganisasian himpunan data ke dalam sejumlah kelompok (*cluster*) spesifik, di mana klasifikasi tersebut ditentukan oleh tingkat kemiripan atribut atau karakteristik inheren yang melekat pada setiap objek data [7]. *K-Means* termasuk ke dalam algoritma *unsupervised learning* yang dirancang untuk mengeksplorasi dan mengidentifikasi pola, struktur, maupun informasi tersembunyi pada data yang tidak memiliki label [8].

Sejumlah kajian terdahulu telah membuktikan reliabilitas algoritma *K-Means* dalam memetakan profil wajib pajak. Sebagai contoh, riset [9] berhasil mengategorikan wajib pajak PBB di Desa Kendal ke dalam empat tingkatan kepatuhan. Serupa dengan hal tersebut, penelitian [10] mengimplementasikan metode yang sama untuk mengklasifikasikan wajib pajak kendaraan bermotor di UPT Samsat Medan Selatan menjadi empat kelompok kedisiplinan. Skalabilitas *K-Means* juga ditunjukkan dalam penelitian [11] yang memetakan penerimaan PBB per kelurahan di Kota Tasikmalaya ke dalam enam *cluster* unik, serta pada studi [12] yang mengelompokkan kepatuhan wajib pajak di Kota Medan berdasarkan data tahun 2023 menjadi tiga tingkatan. Fleksibilitas algoritma ini juga teruji di luar domain perpajakan; misalnya, dalam analisis preferensi konsumen terhadap produk *frozen food* [13] pemetaan sebaran kasus stunting pada balita di Desa Randudongkal [14] hingga segmentasi profil akademik mahasiswa [15]. Ragam studi tersebut menegaskan bahwa *K-Means* merupakan instrumen yang tangguh untuk mengekstraksi informasi dari dataset yang beragam.

Rangkaian studi tersebut membuktikan efektivitas algoritma *K-Means* dalam mengidentifikasi profil serta segmentasi kepatuhan wajib pajak secara akurat. Namun, belum terdapat kajian yang secara spesifik dan mendalam menerapkan algoritma ini pada data wajib pajak PBB-P2 dalam skala besar di

DKI Jakarta, terutama dengan mempertimbangkan aspek perilaku pembayaran dan waktu pembayaran sebagai salah satu indikator kepatuhan.

Sejumlah literatur terdahulu telah membuktikan reliabilitas algoritma *K-Means* dalam memetakan profil wajib pajak melalui tren pembayaran mereka. Meski demikian, implementasi algoritma ini secara komprehensif untuk wajib pajak masuk dalam kategori PBB-P2 di wilayah DKI Jakarta dengan mengintegrasikan metrik kuantitatif berupa volume pembayaran dan disiplin waktu pelunasan masih sangat terbatas, khususnya pada studi yang mengeksplorasi basis data berskala besar dalam rentang waktu lima tahun. Oleh karena itu, riset ini dilakukan guna mengisi celah riset (*research gap*) tersebut dengan mengeksplorasi penggunaan data sekunder mutakhir dalam rentang periode 2020 hingga 2024. Melalui pendekatan *clustering* berbasis perilaku pembayaran, studi ini diharapkan mampu menyumbangkan solusi praktis dalam memformulasikan kebijakan strategis guna meningkatkan kepatuhan pajak daerah.

Temuan studi ini mengonfirmasi bahwa penggunaan teknik *clustering* melalui algoritma *K-Means* sangat berdaya guna dalam mendistribusikan wajib pajak ke dalam beberapa segmen yang menunjukkan kontras tingkat kepatuhan secara eksplisit, baik ditinjau dari aspek nominal setoran maupun kedisiplinan waktu pembayaran. Kategorisasi ini diproyeksikan dapat menyajikan potret komprehensif mengenai realitas perilaku subjek pajak di lapangan, sekaligus menjadi landasan strategis dalam merumuskan kebijakan yang lebih akurat demi mengoptimalkan capaian penerimaan PBB-P2 di area DKI Jakarta.

II. METODE PENELITIAN

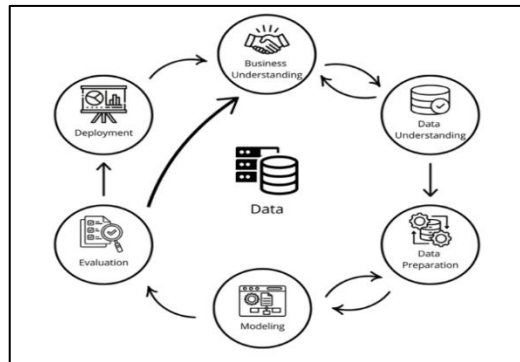
Riset ini menerapkan pendekatan kuantitatif terapan yang difokuskan pada pengategorian wajib pajak PBB-P2 melalui analisis tingkat kedisiplinan mereka dalam memenuhi tanggung jawab fiskal di wilayah Provinsi DKI Jakarta. Metodologi penelitian ini mengadopsi kerangka kerja *Cross-Industry Standard Process for Data Mining* (CRISP-DM), yang mencakup VI fase fundamental: pemahaman bisnis (*business understanding*), pemahaman data (*data understanding*), persiapan data (*data preparation*), pemodelan (*modeling*), evaluasi (*evaluation*), serta penyebaran (*deployment*). Pemilihan pendekatan *CRISP-DM* dilakukan karena sifatnya yang adaptif serta luasnya penerapan dalam berbagai penelitian dan proyek *data mining*. Alur pelaksanaan metode penelitian berdasarkan model *CRISP-DM* ditunjukkan pada Gambar 1.

Business Understanding

Fokus utama dalam riset ini ditujukan pada derajat kedisiplinan wajib pajak PBB-P2 di area DKI Jakarta. Dimensi kepatuhan tersebut menjadi parameter vital dalam tata kelola fiskal daerah lantaran memiliki korelasi linear terhadap optimalisasi Pendapatan Asli Daerah (PAD). Mengingat populasi wajib pajak PBB-P2 di Jakarta telah menyentuh angka kurang lebih 2,2 juta pada tahun 2024, maka pemetaan kepatuhan yang mendalam menjadi urgensi mendasar sebagai basis formulasi kebijakan yang strategis. Penelitian ini bertujuan untuk mengidentifikasi *cluster* wajib pajak berdasarkan tingkat kepatuhan mereka, sehingga pihak pembuat kebijakan dapat merumuskan langkah-langkah strategis untuk meningkatkan kepatuhan pajak. Melalui implementasi pengelompokan tersebut, diharapkan muncul pola-pola signifikan yang dapat dijadikan basis fundamental dalam perumusan kebijakan perpajakan yang cukup objektif, efektif, dan efisien.

Data Understanding

Penelitian ini mengambil data sekunder dari *database* piutang wajib pajak pada PBB-P2 yang dikelola oleh Badan Pendapatan Daerah (*Bapenda*) Provinsi DKI Jakarta. Analisis dilakukan terhadap data selama rentang waktu lima tahun terakhir, yaitu periode 2020 hingga 2024, dengan tujuan untuk mengidentifikasi tren kepatuhan secara longitudinal. Dataset yang dianalisis dalam penelitian ini berjumlah 28.125 entri, yang mencakup informasi Nomor Objek Pajak (NOP), tahun pajak, besaran pajak terutang, jumlah pembayaran yang telah dilakukan, tanggal jatuh tempo, tanggal pelunasan, serta status tunggakan. Contoh data PBB-P2 yang digunakan sebagai bahan analisis dalam penelitian ini ditampilkan pada Gambar 2.



Gambar 1. Metode Penelitian

PBB_NEW	TAHUN_PAJAK	TANGGAL_JATUH_TEMPO	PBB_YG_HARUS_DIBAYAR_SPPT	TGL_LAST_BAYAR	JML_BAYAR_POROK	STATUS
X100020	2020	2020-09-30 00:00:00.000	1697304	NaN	0	BELUM LUNAS
X100020	2021	2021-10-29 00:00:00.000	1697304	NaN	0	BELUM LUNAS
X100020	2022	2022-12-09 00:00:00.000	3138027	NaN	0	BELUM LUNAS
X100020	2023	2023-09-30 00:00:00.000	3312362	NaN	0	BELUM LUNAS
X100020	2024	2024-11-30 00:00:00.000	4132908	NaN	0	BELUM LUNAS
...	...	...	...	...	...	...
I3701990	2020	2020-09-30 00:00:00.000	685145	NaN	0	BELUM LUNAS
I3701990	2021	2021-10-29 00:00:00.000	685145	NaN	0	BELUM LUNAS
I3701990	2022	2022-12-09 00:00:00.000	634378	NaN	0	BELUM LUNAS
I3701990	2023	2023-09-30 00:00:00.000	671695	NaN	0	BELUM LUNAS
I3701990	2024	2024-11-30 00:00:00.000	839618	NaN	0	BELUM LUNAS

Gambar 2. Sampel Data PBB-P2

Dalam pelaksanaan riset ini, terdapat dua atribut kunci yang dijadikan sebagai variabel penelitian, yakni rasio total pembayaran serta tingkat akurasi waktu dalam melakukan pembayaran. Persentase pembayaran diperoleh melalui perbandingan antara akumulasi pajak yang telah dibayarkan dengan total pajak terutang selama periode pengamatan lima tahun. Istilah total pajak terutang merepresentasikan nominal kewajiban finansial yang harus dipenuhi oleh seorang wajib pajak atas objek pajak tertentu dalam periode satu tahun, sementara total pajak yang dibayarkan merujuk pada akumulasi dana yang secara aktual telah disetorkan oleh wajib pajak untuk objek pajak tersebut. Sementara itu, ketepatan waktu pembayaran dihitung berdasarkan proporsi pembayaran yang dilakukan sebelum batas jatuh tempo terhadap jumlah tahun pengamatan, yang selanjutnya dinyatakan dalam bentuk persentase.

Penggunaan atribut persentase pembayaran sebagai fitur dalam penelitian ini didasarkan pada signifikansinya sebagai indikator langsung yang mempengaruhi kepatuhan wajib pajak. Variabel ini mengukur rasio antara total pajak yang dibayarkan terhadap total pajak terutang, yang secara empiris telah terbukti menjadi faktor yang berpengaruh dalam menilai tingkat kepatuhan. Sedangkan atribut persentase waktu pembayaran menunjukkan bahwa ketepatan waktu pembayaran tidak hanya mencerminkan kepatuhan, tetapi juga tingkat kesadaran dan disiplin wajib pajak. Melalui perbandingan antara frekuensi pelunasan sebelum jatuh tempo dengan total rentang waktu observasi, variabel ini mampu mengartikulasikan pola pembayaran secara lebih komprehensif [16].

Data Preparation

Tahap ini dilakukan serangkaian prosedur prapemrosesan data yang meliputi eliminasi entitas data yang tidak konsisten (*data cleaning*), transformasi variabel, serta verifikasi akhir guna memastikan bahwa dataset telah memenuhi standar kelayakan untuk dieksekusi lebih lanjut menggunakan algoritma yang relevan. Data ditransformasi dengan melakukan normalisasi skala menggunakan metode *StandardScaler* agar seluruh atribut memiliki nilai yang sebanding. Hasil dari tahap ini akan digunakan untuk proses *modeling*.

Modeling

Tahapan segmentasi dilakukan dengan mengintegrasikan algoritma *K-Means*, sebuah teknik yang berfokus pada partisi objek data ke dalam kluster-kluster spesifik. Mekanisme ini bekerja melalui penghitungan jarak Euclidean minimum untuk mengatribusikan setiap titik data ke *centroid* atau titik

pusat massa kelompok yang paling relevan. Melalui mekanisme tersebut, entitas data yang menunjukkan keseragaman karakteristik akan diklasifikasikan ke dalam satu kategori yang homogen, sementara data yang memiliki profil divergen akan secara sistematis dipisahkan ke dalam kelompok yang berbeda [17]. Jarak antar data dan *centroid* dihitung dengan mengadopsi rumus *Euclidean Distance* yang diberikan pada persamaan (1).

$$d(x, y) = \sqrt{\sum_i^n (x_i - y_i)^2}, \quad (1)$$

$d(x, y)$ adalah jarak antara titik data x_i dan *centroid* y_i , dan x_i adalah nilai atribut ke- i dari titik data. Jarak terkecil menentukan pengelompokan data ke dalam *cluster* tertentu.

Untuk menetapkan kuantitas kelompok yang paling ideal, penelitian ini menerapkan teknik *Elbow Method* dengan melakukan observasi terhadap tren penurunan angka *inertia*. Pendekatan *Elbow* ini berfungsi untuk mengidentifikasi jumlah *cluster* terbaik melalui evaluasi korelasi antara nilai K (jumlah kelompok) dengan skor yang dihasilkan. Penentuan titik optimal dilakukan dengan mengidentifikasi lekukan tajam pada grafik yang menyerupai bentuk siku (*elbow*), di mana titik tersebut merepresentasikan jumlah kelompok yang paling ideal dalam menggambarkan struktur data [18]. Setelah jumlah kluster ditetapkan, algoritma akan menjalankan proses iterasi secara repetitif hingga posisi *centroid* mencapai titik ekuilibrium (stabil), di mana tidak ditemukan lagi pergeseran atau relokasi objek data antar kelompok.

Evaluation

Evaluasi hasil *clustering* dilakukan dengan menggunakan *Silhouette Analysis*. Teknik ini mengevaluasi validitas posisi setiap objek dalam suatu kluster dengan mengukur tingkat kohesi terhadap anggota di kelompok yang sama serta tingkat separasi atau jaraknya terhadap anggota di kelompok yang berbeda [19]. Nilai koefisien *silhouette* yang mendekati 1 menunjukkan kualitas pengelompokan yang kuat, sedangkan nilai yang kurang dari sama dengan 0.25 menunjukkan kualitas pengelompokan yang tidak terstruktur [20].

Deployment

Tahapan ini difokuskan pada perumusan laporan akhir riset dengan tujuan agar seluruh temuan dapat dipahami secara lebih komprehensif. Selain itu, dokumentasi ini berfungsi sebagai rujukan strategis dalam membedah pola perilaku para wajib pajak PBB-P2 di area DKI Jakarta, serta menjadi landasan dalam merumuskan kebijakan yang lebih akurat guna menstimulasi peningkatan kepatuhan perpajakan.

III. HASIL DAN PEMBAHASAN

Dataset yang dianalisis dalam penelitian ini mencakup catatan pembayaran PBB-P2 dari 28.125 wajib pajak di wilayah Provinsi DKI Jakarta, dengan rentang observasi data selama periode tahun 2020 hingga 2024. Dari data mentah yang mencakup nominal terutang dan nominal pembayaran per tahun, dilakukan perhitungan persentase pembayaran untuk masing-masing objek pajak. Persentase ini menunjukkan seberapa besar kewajiban pajak yang telah dipenuhi oleh wajib pajak dalam lima tahun terakhir. Sampel data persentase pembayaran dapat dilihat pada Tabel 1.

Di samping itu, proses pengolahan data turut dilakukan terhadap variabel waktu pembayaran sebagaimana tercantum pada Tabel 2. Parameter ketepatan waktu tersebut diestimasi dengan membandingkan kesesuaian antara tanggal realisasi penyeteroran pajak dengan batas waktu jatuh tempo yang telah ditetapkan. Evaluasi dilakukan secara berkala pada tiap tahunnya untuk mengukur kedisiplinan pembayaran. Output akhir dari proses ini direpresentasikan dalam bentuk persentase ketepatan waktu selama rentang observasi lima tahun. Kedua atribut tersebut yakni rasio pelunasan pajak dan persentase ketepatan waktu menjadi basis data fundamental dalam prosedur segmentasi menggunakan algoritma *K-Means*. Proses pengolahan data dilakukan menggunakan bahasa pemrograman Python. Setelah data siap, dilakukan standarisasi menggunakan transformasi skala dari *library scikit-learn* agar kedua variabel memiliki bobot yang seimbang dalam proses analisis. Hasil standarisasi dapat dilihat pada Tabel 3.

Selanjutnya, pemilihan jumlah *cluster* yang paling tepat dilakukan dengan menggunakan metode *Elbow*, yang menampilkan grafik penurunan nilai *inertia* seiring bertambahnya jumlah *cluster*. Titik yang dinilai paling ideal teridentifikasi saat reduksi angka *inertia* mulai menunjukkan perlambatan yang signifikan, sehingga menciptakan lekukan yang menyerupai bentuk 'siku' pada visualisasi grafik. Representasi dari implementasi teknik *Elbow* tersebut dapat dicermati pada Gambar 3.

Tabel 1. Sampel Data Persentase Pembayaran

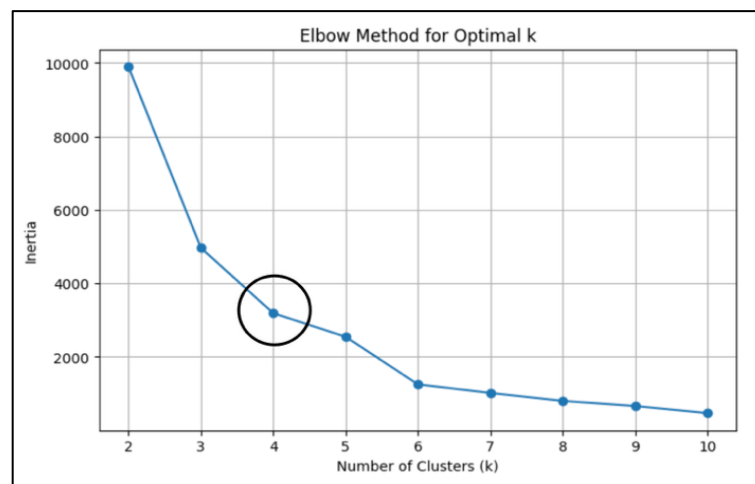
Nomor Objek Pajak	Total Jumlah Pajak Terutang (Rupiah)	Total Jumlah Pajak Dibayar (Rupiah)	Persentase Pembayaran (%)
317XXXXXXX360	13.068.330	13.068.330	100,0
317XXXXXXX560	637.485	18.420	2,89
317XXXXXXX730	74.230	59.301	79,89

Tabel 2. Sampel Data Persentase Pembayaran Tepat Waktu

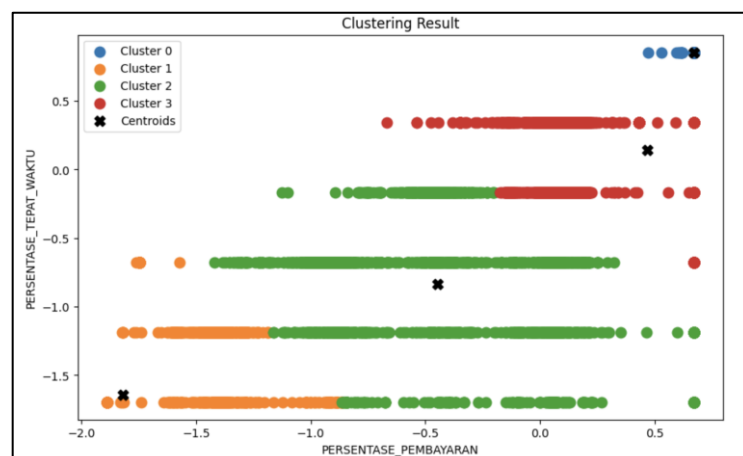
Nomor Objek Pajak	Jumlah Pembayaran Tepat Waktu	Persentase Pembayaran Tepat Waktu (%)
317XXXXXXX360	1	20,0
317XXXXXXX560	0	0,0
317XXXXXXX730	1	20,0

Tabel 3. Data Hasil Standarisasi

Nomor Objek Pajak	Persentase Pembayaran	Persentase Pembayaran Tepat Waktu
317XXXXXXX360	0,6754639543812325	-1,043413355356199
317XXXXXXX560	-1,6642672067830449	-1,528260621547695
317XXXXXXX730	0,9094131355110572	-1,043413355356199

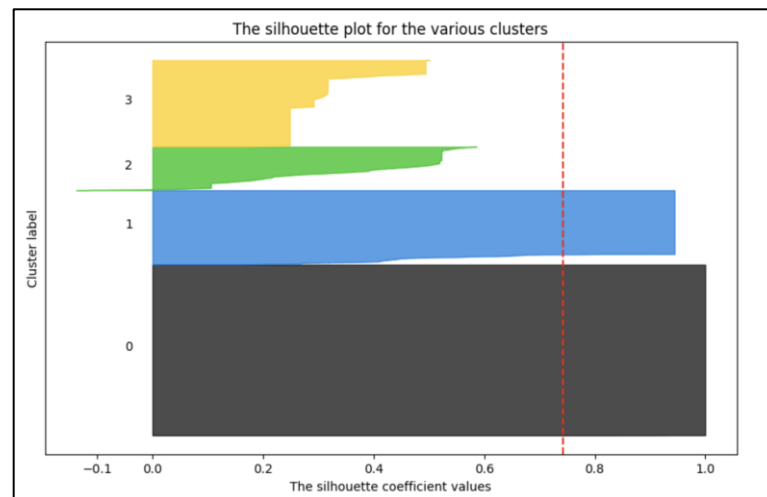


Gambar 3. Grafik *Elbow Method*



Gambar 4. Visualisasi Hasil Pengelompokan

Tabel 4. Hasil Pengelompokan Wajib Pajak		
Cluster	Jumlah Wajib Pajak	Kategori Kepatuhan
0	12.806	Sangat Tinggi
1	5.578	Rendah
2	3.252	Sedang
3	6.489	Tinggi



Gambar 5. Hasil Evaluasi *Silhouette Analysis*

Tabel 5. Karakteristik Setiap Cluster			
Cluster	Rata-Rata Persentase Pembayaran (%)	Rata-Rata Tepat Waktu (%)	Karakteristik
0	99,99	100,0	Selalu membayar penuh dan tepat waktu
1	2,68	2,11	Sangat rendah dalam pembayaran dan ketepatan waktu
2	56,30	33,69	Cukup dalam pembayaran, namun sering terlambat
3	92,04	72,11	Hampir seluruhnya dibayar dan mayoritas tepat waktu

Dari hasil analisis tersebut, ditetapkan bahwa jumlah *cluster* optimal adalah empat, dengan alasan adanya titik siku (*elbow*) pada grafik di *cluster* keempat. Model *K-Means* kemudian diinisialisasi dengan parameter jumlah *cluster* = 4. Proses pelatihan dilakukan hingga *centroid* setiap *cluster* stabil. Setiap wajib pajak dalam *dataset* mendapatkan label *cluster* berdasarkan kedekatannya dengan *centroid* yang paling sesuai. Visualisasi persebaran data wajib pajak PBB-P2 setelah dilakukan pengelompokan dapat dijelaskan pada Gambar 4 dan kategori hasil pengelompokan dapat dilihat pada Tabel 4.

Merujuk pada data yang disajikan dalam Tabel 4, mayoritas wajib pajak masuk ke dalam kategori dengan tingkat kepatuhan sangat maksimal (*Cluster* 0), sementara populasi terkecil ditemukan pada kelompok dengan tingkat kepatuhan moderat (*Cluster* 2). Fenomena tersebut merefleksikan bahwa mayoritas wajib pajak memiliki preferensi terhadap perilaku disiplin dalam mengeskalasi pemenuhan kewajiban PBB-P2 mereka, walaupun masih terdapat segmen dengan tingkat kepatuhan rendah yang memerlukan penanganan lebih spesifik. Tahap berikutnya adalah melakukan evaluasi guna menguji efektivitas model dalam menjawab permasalahan yang telah dirumuskan. Pada riset *clustering* ini, proses validasi dilakukan melalui *silhouette analysis* untuk mengukur derajat kualitas dari kelompok-kelompok yang telah dihasilkan.

Visualisasi pada Gambar 5 menampilkan sumbu X yang merujuk pada skor *silhouette coefficient* bagi tiap poin data, yang terbentang dalam skala antara -1 hingga 1. Perolehan angka yang kian mendekati 1 memberikan sinyal bahwa posisi data tersebut sangat jauh dari kelompok lain, sehingga dapat dikatakan telah terlokasi dengan sangat tepat pada *cluster*-nya. Sebaliknya, skor yang berada di sekitar angka 0 mencerminkan bahwa data tersebut terletak pada zona transisi atau perbatasan antar dua kelompok, sementara nilai di bawah nol (negatif) memberikan indikasi bahwa data kemungkinan besar mengalami salah penempatan dan lebih cocok dikategorikan ke dalam kelompok yang berbeda. Sumbu Y pada Gambar 5 menampilkan label *cluster*, di mana setiap baris horizontal mewakili *cluster* yang berbeda.

Indikator garis putus-putus merah secara vertikal merepresentasikan nilai rata-rata *Silhouette Coefficient* dari keseluruhan dataset. Parameter ini berfungsi sebagai metrik evaluasi global yang menggambarkan efektivitas pengelompokan data, di mana peningkatan nilai rata-rata berbanding lurus dengan optimalnya kualitas klusterisasi yang dihasilkan.

Eksperimen klusterisasi memberikan nilai rata-rata *Silhouette Coefficient* sebesar 0,742. Skor tersebut menjadi bukti empiris bahwa skema pengelompokan yang dihasilkan memiliki struktur yang sangat kuat, valid, dan menunjukkan pemisahan antarkelompok yang sangat jelas, ditandai dengan pemisahan jarak antar *cluster* yang tegas serta derajat kepadatan objek data yang sangat tinggi di dalam setiap kelompoknya. Hasil ini menegaskan bahwa penerapan algoritma *K-Means* pada data wajib pajak PBB-P2 mampu mengidentifikasi kelompok kepatuhan secara terstruktur, sehingga berpotensi menjadi alat bantu pengambilan keputusan yang efektif. Berdasarkan hasil pengelompokan tersebut, dilakukan analisis deskriptif untuk memahami karakteristik masing-masing *cluster*. Rangkuman dari analisis ini dapat dilihat pada Tabel 5.

Sesuai dengan Tabel 5, kelompok-kelompok tingkat kepatuhan yang terdapat pada empat *cluster* yang berbeda dapat diuraikan strategi kebijakan dan manfaat yang dapat diperoleh oleh pemangku kebijakan sebagai berikut. *Cluster* 0, yang mewakili kelompok wajib pajak dengan tingkat kepatuhan sangat tinggi, ditandai oleh rata-rata persentase pembayaran sebesar 99,99% serta ketepatan waktu pembayaran mencapai 100%. Kelompok ini menunjukkan tingkat disiplin yang konsisten dalam melaksanakan kewajiban perpajakan mereka. Oleh sebab itu, strategi kebijakan yang representatif bagi kelompok ini adalah dengan memitigasi penurunan kepatuhan melalui penguatan motivasi dan loyalitas, salah satunya melalui skema apresiasi berupa pemberian penghargaan atau insentif khusus.

Sebaliknya, *cluster* 1 merepresentasikan kelompok wajib pajak dengan kepatuhan yang sangat rendah, dengan rata-rata persentase pembayaran hanya 2,68% serta ketepatan waktu pembayaran yang sangat buruk. Kondisi ini mengindikasikan adanya kendala signifikan baik dari segi pemahaman maupun kesadaran wajib pajak terhadap kewajiban perpajakan. Oleh karena itu, pemangku kebijakan perlu melakukan intervensi yang intensif melalui program edukasi perpajakan sekaligus memperkuat penegakan hukum. Penerapan sanksi administratif yang tegas serta mekanisme pengawasan dan penagihan yang efektif diharapkan dapat mengurangi tingkat ketidakpatuhan, sehingga meningkatkan basis penerimaan wajib pajak.

Pada *cluster* 2, karakteristik yang muncul menunjukkan tingkat pembayaran yang moderat sebesar 56,30% namun dengan tingkat ketepatan waktu yang relatif rendah, yaitu 33,69%. Hal ini menunjukkan adanya kemampuan membayar pada tingkat yang cukup baik, tetapi dengan disiplin waktu yang kurang optimal. Oleh sebab itu, strategi yang dapat diimplementasikan adalah melalui pengembangan sistem pengingat pembayaran secara berkala serta penyesuaian kebijakan insentif yang dapat mendorong perilaku tepat waktu. Dengan upaya ini, penerimaan pajak menjadi lebih stabil dan dapat diprediksi, sehingga dapat meningkatkan tingkat penerimaan pajak.

Pada *cluster* 3, menunjukkan tingkat kepatuhan yang relatif tinggi dengan rata-rata pembayaran 92,04% dan ketepatan waktu yang cukup baik sebesar 72,11%. Dalam hal ini, kebijakan yang difokuskan pada penyederhanaan prosedur administrasi perpajakan dan meningkatkan komunikasi dengan wajib pajak sangat diperlukan guna mempermudah pelaksanaan kewajiban perpajakan. Implementasi pendekatan ini tidak hanya akan mengeskalasi optimalisasi penerimaan pajak, tetapi juga memperkokoh kepuasan serta loyalitas wajib pajak. Hal tersebut pada gilirannya akan memberikan dampak positif dalam menjaga stabilitas kepatuhan dalam jangka panjang.

Secara holistik, pendalaman yang menyeluruh terhadap karakteristik spesifik dari setiap segmen kepatuhan ini memberikan peluang bagi para pengambil keputusan untuk memformulasikan kebijakan

yang lebih terspesialisasi, berdaya guna, serta efisien dalam penggunaan sumber daya. Dengan demikian, penerapan kebijakan yang berbasis data karakteristik wajib pajak akan berkontribusi signifikan dalam meningkatkan kepatuhan secara menyeluruh, sekaligus mendukung tujuan pembangunan daerah melalui peningkatan penerimaan pajak yang berkelanjutan.

IV. KESIMPULAN

Penelitian ini bertujuan mengelompokkan wajib pajak *PBB-P2* di Provinsi DKI Jakarta berdasarkan tingkat kepatuhan mereka dalam membayar pajak dengan menggunakan algoritma *K-Means*. Analisis terhadap 28.125 data wajib pajak dari periode 2020 hingga 2024 menunjukkan terbentuknya empat kelompok kepatuhan yang berbeda. Kelompok pertama menunjukkan tingkat kepatuhan yang sangat tinggi dengan konsistensi pembayaran penuh dan tepat waktu, sedangkan kelompok lainnya menunjukkan variasi tingkat kepatuhan dari rendah hingga tinggi.

Perolehan angka rata-rata *silhouette coefficient* yang mencapai 0,742 membuktikan bahwa model *clustering* yang dikembangkan memiliki mutu segmentasi yang unggul. Hal ini menandakan bahwa tiap kelompok bentukan memiliki konsistensi karakteristik internal yang kuat serta diferensiasi yang tegas terhadap kelompok lainnya. Temuan ini mengonfirmasi efektivitas implementasi teknik *data mining* melalui algoritma *K-Means* dalam mengonstruksi pola perilaku wajib pajak secara mendalam. Lebih jauh lagi, pemetaan tersebut menyediakan profil kepatuhan *PBB-P2* yang komprehensif di Jakarta, yang dapat diutilisasi sebagai basis fundamental dalam perumusan kebijakan strategis guna memaksimalkan potensi pendapatan fiskal daerah.

Pada pengembangan penelitian di masa mendatang, disarankan untuk menambahkan variabel-variabel eksternal yang lebih kompleks agar analisis menjadi lebih komprehensif. Penelitian selanjutnya juga dapat membandingkan hasil *clustering* dengan metode lain untuk mengevaluasi keandalan metode yang digunakan. Selain itu, pengembangan visualisasi dalam bentuk *dashboard* interaktif dapat memperluas pemanfaatan hasil penelitian ini dalam mendukung pengambilan keputusan oleh pemerintah daerah dengan profesional dan tepat sasaran.

DAFTAR PUSTAKA

- [1] A. N. Nabawi, K. I. Absy, S. R. Dalimunthe, dan D. Vientiany, "Pajak Pusat dan Pajak Daerah," vol. 02, no. 01, hlm. 1737–1744, 2025.
- [2] Amelia Kamlannabila, Syamsul Huda, dan Putra Perdana, "Optimalisasi Pembayaran Pajak Bumi dan Bangunan melalui Penyampaian Informasi Tagihan Secara Door-to-Door di Kecamatan Sambikerep," *GERVASI: Jurnal Pengabdian kepada Masyarakat*, 2025.
- [3] A. Shabrina Ilra dan M. Aprilya, "Pajak Bumi dan Bangunan PBB Land and Building Tax," *JIIIC: JURNAL INTELEK INSAN CENDIKIA*, Jun 2025, [Daring]. Tersedia pada: <https://jicnusantara.com/index.php/jiic>
- [4] D. M. Parapat, "Faktor yang Mempengaruhi Kepatuhan Wajib Pajak Membayar Pajak Bumi dan Bangunan di DKI Jakarta," 2023.
- [5] Angelina Jessica Hirawan dan Widyasari, "Faktor yang Mempengaruhi Kepatuhan Wajib Pajak dalam Membayar PBB Daerah Jakarta Barat," 2021.
- [6] E. Mayoana Fitri, R. Randy Suryono, dan A. Wantoro, "Klasterisasi Data Penjualan Berdasarkan Wilayah Menggunakan Metode K-Means pada PT XYZ."
- [7] L. Y. Hutabarat, I. Gunawan, I. Purnamasari, M. Safii, dan W. Saputra, "Penerapan Algoritma K-Means dalam Pengelompokan Jumlah Penduduk Berdasarkan Kelurahan di Kota Pematangsiantar," *Jurnal Ilmu Komputer dan Teknologi*, vol. 2, no. 2, hlm. 20–26, 2021, [Daring]. Tersedia pada: <http://creativecommons.org/licenses/by/4.0/>
- [8] I. H. Sarker, "AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems," *SN Comput Sci*, vol. 3, no. 2, Mar 2022, doi: 10.1007/s42979-022-01043-x.
- [9] M. A. Putri, N. Rahaningsih, F. M. Basysyar, dan O. Nurdiawan, "Penerapan Data Mining Menggunakan Metode Clustering Untuk Mengetahui Kelompok Kepatuhan Wajib Pajak Bumi dan Bangunan," *Jurnal Accounting Information System (AIMS)*, vol. 5, no. 2, hlm. 145–156, 2022, doi: 10.32627.

- [10] D. Asti, M. S. Hasibuan, dan P. A. Siregar, “Penerapan Algoritma K-Means untuk Mengetahui Tingkat Kepatuhan Wajib Pajak Kendaraan Bermotor pada UPT Samsat Medan Selatan,” *Journal of Computer Science and Informatics Engineering (CoSIE)*, 2023.
- [11] R. Nursaniah, N. Rahaningsih, I. Ali, dan N. Dienwati Nuris, “Pengelompokan Data Penerimaan Pajak Bumi dan Bangunan Berdasarkan Kelurahan di Kota Tasikmalaya Menggunakan Algoritma K-Means,” 2024.
- [12] Sri Wahyuni dan Sriani, “Penerapan Algoritma K-Means untuk Pengelompokan Kepatuhan Wajib Pajak Bumi dan Bangunan di Kota Medan,” 2025. [Daring]. Tersedia pada: <https://jurnal.unimed.ac.id/2012/index.php/cess>
- [13] T. Amalina, D. Bima, A. Pramana, dan B. N. Sari, “Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food,” *Jurnal Ilmiah Wahana Pendidikan*, vol. 8, no. 15, hlm. 574–583, 2022, doi: 10.5281/zenodo.7052276.
- [14] R. N. Pratistha dan B. Kristianto, “Implementasi Algoritma K-Means dalam Klasterisasi Kasus Stunting pada Balita di Desa Randudongkal,” 2024. [Daring]. Tersedia pada: <https://journal.stmiki.ac.id>
- [15] Y. Suhandi, I. Kurniati, dan S. Norma, “Penerapan Metode CRIPS-DM dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik,” *Jurnal Teknologi Informatika dan Komputer*, vol. 6, no. 2, hlm. 12–20, Sep 2020, doi: 10.37012/jtik.v6i2.299.
- [16] Mohd Rizal Palil, “Factors affecting tax compliance behaviour in self assessment system,” *AFRICAN JOURNAL OF BUSINESS MANAGEMENT*, vol. 5, no. 33, Des 2011, doi: 10.5897/ajbm11.1742.
- [17] A. N. Khasanah, V. A. Tricahyo, dan M. M. Huda, “Implementasi Metode K-Means Clustering Sebagai Penentu Kelompok Belajar di SMA,” *Journal of Information System Research*, vol. 6, no. 2, hlm. 1281–1289, 2025, doi: 10.47065/josh.v6i2.6631.
- [18] A. E. Putri, Y. Mz, dan J. E. Bororing, “Implementasi K-Means Clustering dan Model CRIPS-DM untuk Pengelompokan Daerah Rawan Tindak Pidana Narkoba di DIY,” *Jurnal Informatika Teknologi dan Sains*, vol. 7, no. 2, hlm. 615, 2025.
- [19] F. N. Dhewayani, D. Amelia, D. N. Alifah, B. N. Sari, dan M. Jajuli, “Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM,” *Jurnal Teknologi dan Informasi*, 2022, doi: 10.34010/jati.v12i1.
- [20] R. Hidayati, A. Zubair, A. Hidayat Pratama, L. Indana, P. Studi Sistem Informasi, dan F. Teknologi Informasi, “Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering Silhouette Coefficient Analysis in 6 Measuring Distances of K-Means Clustering,” 2021.