Pengelompokan Perangkat Daerah di Jawa Barat Berdasarkan Jenis Pajak Menggunakan Metode *K-Medoids*

Alma Zahra Sartono *, Abdul Kudus

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

almazahrasartono@gmail.com, abdul.kudus@unisba.ac.id

Abstract. The management of regional taxes is one of the critical aspects of increasing locally generated revenue (PAD) that supports regional development. In West Java, with a large and diverse number of regional agencies, an effective strategy is needed to understand the characteristics of each agency in terms of contributions and the types of taxes managed. One approach that can be utilized is grouping regional agencies based on tax types using cluster analysis. Among the many clustering algorithms available, the K-Medoids algorithm stands out due to its robustness against outliers. This method is considered highly suitable as tax data often contains extreme values that can influence clustering results. Based on the clustering analysis conducted, two clusters were formed: Cluster 1 consists of 1 regional agency, PD1, while Cluster 2 includes 37 regional agencies. The average value for each type of tax in Cluster 1 is higher than in Cluster 2, indicating that members of Cluster 1 are high-tax contributors, whereas members of Cluster 2 are low-tax contributors.

Keywords: *Cluster Analysis, K-Medoids, Regional Agency, Tax.*

Abstrak. Pengelolaan pajak daerah merupakan salah satu aspek penting dalam meningkatkan pendapatan asli daerah (PAD) yang mendukung pembangunan wilayah. Di Jawa Barat, dengan jumlah perangkat daerah yang tidak sedikit dan beragam, diperlukan strategi yang efektif untuk memahami karakteristik masing-masing perangkat daerah dalam hal kontribusi dan jenis pajak yang dikelola. Salah satu pendekatan yang dapat digunakan adalah pengelompokan perangkat daerah berdasarkan jenis pajak menggunakan analisis klaster. Dalam konteks pengelompokan perangkat daerah, di antara banyaknya algoritma analisis klaster, algoritma K-Medoids menarik perhatian karena sifatnya yang lebih tahan terhadap pencilan. Metode ini dirasa sangat cocok karena data pajak sering kali mengandung nilai ekstrim yang dapat memengaruhi hasil klasterisasi. Berdasarkan hasil pengelompokan yang telah dilakukan, diperoleh sebanyak 2 klaster dengan jumlah anggota klaster 1 sebanyak 1 perangkat daerah yaitu PD1 dan anggota klaster 2 sebanyak 37 perangkat daerah. Nilai rata-rata tiap jenis pajak pada klaster 1 lebih tinggi dibandingkan dengan klaster 2, sehingga dapat dikatakan bahwa anggota klaster 1 merupakan anggota klaster dengan tarikan pajak yang tinggi, sedangkan anggota klaster 2 merupakan anggota klaster dengan tarikan pajak yang rendah.

Kata Kunci: *Analisis Klaster, K-Medoids, Perangkat Daerah, Pajak.*

*abdul.kudus@unisba.ac.id

A. Pendahuluan

Pengelolaan pajak daerah merupakan salah satu aspek penting dalam meningkatkan pendapatan asli daerah (PAD) yang mendukung pembangunan wilayah. Di Jawa Barat, dengan jumlah perangkat daerah yang tidak sedikit dan beragam, diperlukan strategi yang efektif untuk memahami karakteristik masing-masing perangkat daerah dalam hal kontribusi dan jenis pajak yang dikelola. Salah satu pendekatan yang dapat digunakan adalah pengelompokan perangkat daerah berdasarkan jenis pajak.

Pengelompokan ini dapat memberikan wawasan penting bagi pemerintah provinsi Jawa Barat dalam merancang kebijakan yang lebih terarah, seperti penyusunan program pendampingan atau optimalisasi potensi pajak. Dalam hal ini, analisis klaster menjadi metode yang sangat relevan untuk mengelompokan perangkat daerah berdasarkan jenis pajak. Analisis klaster merupakan suatu metode yang tepat untuk mengklasifikasikan beberapa objek dengan karateristik yang sama (Hair et al., 2006 dalam Widyadhana et al., 2021). Pendekatan ini memungkinkan identifikasi pola tersembunyi dalam dataset tanpa memerlukan informasi label sebelumnya.

Secara garis besar metode dalam analisis klaster terbagi menjadi dua yaitu metode hierarki dan non hierarki. Masing-masing metode memiliki karateristik, keunggulan, dan keterbatasannya tersendiri (Nurfauzan & Darwis, 2024). Metode yang paling banyak digunakan adalah analisis klaster non hierarki. Analisis klaster dengan metode non hierarki atau dapat disebut dengan pengklasteran berbasis *partitioning* merupakan suatu metode yang digunakan untuk menghasilkan partisi dari sebuah data. Menurut Triyanto (2015) hal tersebut akan membuat suatu objek atau karateristik dalam satu klaster menjadi lebih mirip satu dengan yang lainnya. Analisis klaster dengan metode non hierarki dilakukan dengan menentukan jumlah klasternya terlebih dahulu dan tidak dilakukan secara bertahap (Machfudhoh dan Wahyuningsih, 2013).

Dalam konteks pengelompokan perangkat daerah, di antara banyaknya algoritma analisis klaster non hierarki, analisis klaster dengan algoritma *K-Medoids* menarik perhatian karena sifatnya yang lebih tahan terhadap pencilan. Metode ini dirasa sangat cocok karena data pajak sering kali mengandung nilai ekstrim yang dapat memengaruhi hasil klasterisasi. Algoritma *K-Medoids* bekerja dengan memilih objek aktual sebagai pusat klaster (*medoid*) yang kemudian dihitung jarak kemiripan antara *medoid* dengan objek *non-medoid* menggunakan rumus jarak *euclidean*. Menurut Vercellis (2009) dalam Nurholik (2022), hal tersebut bertujuan agar dapat meminimalisir sensitivitas dari partisi dengan nilai yang tinggi pada dataset.

Berdasarkan data yang ada dan latar belakang yang telah diuraikan, penulis ingin melakukan pengelompokan untuk seluruh perangkat daerah yang ada di Provinsi Jawa Barat menjadi beberapa klaster. Hal ini dilakukan agar identifikasi karakteristik tiap perangkat daerah lebih jelas, rinci, dan mudah dibandingkan berdasarkan tarikan pajak tertinggi dan terendah.

Penelitian terkait pengelompokan perangkat daerah berdasarkan jenis pajak dengan menggunakan algoritma *K-Medoids* masih terbatas, khususnya di Provinsi Jawa Barat. Oleh karena itu, penelitian ini bertujuan untuk menerapkan algoritma *K-Medoids* dalam mengelompokan perangkat daerah di Jawa Barat berdasarkan jenis pajak. Penelitian ini diharapkan dapat memberikan kontribusi dalam mendukung pengambilan kebijakan berbasis data yang lebih efektif dan efisien. Kemudian, diharapkan dari penelitian ini dapat memberikan berbagai manfaat bagi penulis maupun pembaca sebagai informasi tambahan mengenai jenis perpajakan dan menambah wawasan mengenai analisis klaster terutam untuk analisis klaster dengan algoritma *K-Medoids*.

B. Metode

Peneliti menggunakan metode analisis klaster dengan algoritma *K-Medoids*. Populasi yang dipilih dalam penelitian ini adalah seluruh perangkat daerah yang ada di Provinsi Jawa Barat dengan jumlah sebanyak 38 perangkat daerah.

Data pada penelitian ini merupakan data sekunder yang diperoleh dari instansi terkait. Data yang digunakan merupakan data tarikan pajak di Sistem Informasi Pemerintah Daerah (SIPD) pada perangkat daerah di Jawa Barat bulan Januari-Agustus 2023. Unit pengamatan dari penelitian ini adalah perangkat daerah yang terdiri dari 38 perangkat daerah. Variabel yang digunakan pada penelitian ini merupakan jenis pajak yang terdiri dari Pajak Pertambahan Nilai (PPN), Pajak Penghasilan Pasal 21 (PPh 21), Pajak Penghasilan Pasal 22 (PPh 22), Pajak Penghasilan Pasal 23 (PPh 23), dan Pajak Penghasilan Pasal 4 ayat (2).

Adapun tahapan dari algoritma *K-Medoids* adalah sebagai berikut:

1. Menentukan *K* (banyaknya klaster) yang ingin dibentuk, yakni $K = 2$.
2. Pilih secara acak sebanyak *K* dari *n* data sebagai pusat klaster (*medoid*) awal.
3. Hitung jarak masing-masing objek *non-medoid* ke *medoid* awal pada setiap klaster menggunakan persamaan jarak *euclidean* dengan rumus sebagai berikut:

$$d_{ij} = \sqrt{\sum_{k=1}^{p}(x_{ik} - x_{jk})^2}\ ,\ k = 1,2,\dots,p \quad \dots (1)$$

Keterangan:

d_{ij} : Jarak *euclidean* dari objek ke-*i* dan objek ke-*j*.
p : Jumlah variabel klaster.
x_{ik} : Nilai objek ke-*i* pada variabel ke-*k*.
x_{jk} : Nilai objek ke-*j* pada variabel ke-*k*.

4. Tandai jarak terdekat objek *non-medoid* ke *medoid*, kemudian hitung total jaraknya.
5. Menentukan anggota klaster terhadap *medoid* awal dengan melihat nilai minimumnya.
6. Lakukan iterasi dan ulangi langkah (1) sampai (5) dengan menggunakan *medoid* baru.
7. Hitung total simpangan (*S*) dengan cara mencari selisih antara total jarak baru dengan total jarak awal. Jika diperoleh nilai *S > 0* maka proses *clustering* berhenti dan hasil klaster adalah pada percobaan pertama. Akan tetapi, jika nilai *S < 0* maka proses iterasi berlanjut.
8. Naikkan *K* menjadi *K + 1*, kemudian kerjakan kembali langkah (2) sampai dengan (7) untuk *K = 2, 3, 4,* dan *5*.
9. Menentukan jumlah klaster optimalnya dengan menggunakan validasi klaster metode indeks *Dunn*, indeks *Connectivity*, dan *Silhouette Coefficient* dengan masing-masing formulanya adalah sebagai berikut:

Indeks *Dunn*

$$DI(c) = min_{1\le i\le n}\left\{min_{1\le j\le n}\left\{\frac{d(C_i,C_j)}{max_{1\le k\le n}\{d'(C_k)\}}\right\}\right\} \quad \dots (2)$$

Keterangan:

$i \neq j$: Pengamatan ke-i tidak boleh sama dengan pengamatan ke-j.
$d(C_i, C_j)$: Jarak antar klaster C_i dan C_j.
$d'(C_k)$: Jarak dalam klaster C_k.
n : banyaknya jumlah klaster.

Nilai terbesar dari indeks Dunn ditetapkan sebagai jumlah optimum klaster (Azuaje, 2001 dalam Paramadina, dkk., 2019). Dengan demikian, nilai indeks Dunn terbesar menunjukkan klaster yang terbaik.

Indeks *Connectivity*

Indeks ini memiliki nilai dimulai dari 0 sampai tak hingga. Semakin kecil nilai indeks *Connectivity*, semakin baik klaster yang terbentuk (Natasya Afira dan Arie Wahyu Wijayanto, 2019).

$$Conn(C) = \sum_{i=1}^{N}\sum_{j=1}^{L} X_{i,nn_{i(j)}} \quad \dots (3)$$

Keterangan:

L : Parameter yang menentukan jumlah tetangga yang berkontribusi untuk nilai indeks.
$nn_{i(j)}$: Pengamatan tetangga terdekat *i* ke-*j*.

Silhouette Coefficient

Indeks ini mengukur derajat kepercayaan dalam proses pengklasteran yang mempunyai kriteria klaster dikatakan terbentuk baik ketika nilai indeks mendekati 1, begitupun kondisi sebaliknya jika nilai indeks mendekati -1 (Halim dan Widodo, 2017 dalam Afira dan Wijayanto, 2021).

$$SI_i^j = \frac{b_i^j - a_i^j}{\max\{a_i^j, b_i^j\}} \quad \text{... (4)}$$

Keterangan:

SI_i^j : Indeks *Silhouette* data ke-*i* dalam satu klaster.

a_i^j : Jarak rata-rata antara pengamatan ke-*i* dan seluruh pengamatan lainnya pada klaster yang sama.

b_i^j : Jarak rata-rata antara pengamatan ke-*i* dan seluruh pengamatan lainnya pada klaster terdekat.

Kemudian, dihitung rata-rata indeks *silhouette* dalam satu klaster dengan menggunakan persamaan berikut:

$$SI_j = \frac{1}{m_j}\sum_{i=1}^{m_j} SI_i^j \quad \text{... (5)}$$

Keterangan:

SI_j : Rata-rata indeks *Silhouette* klaster ke-*j*.

SI_i^j : Indeks *Silhouette* data ke-*i* dalam satu klaster.

m_j : Jumlah data dalam klaster ke-*j*.

Berikut merupakan rumus untuk menghitung nilai *SI* global:

$$SI = \frac{1}{K}\sum_{j=1}^{K} SI_j \quad \text{... (6)}$$

Keterangan:

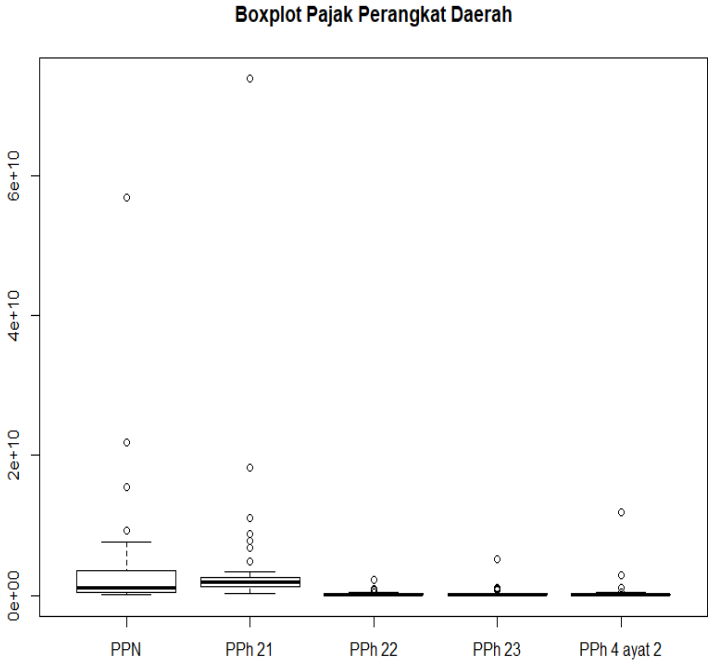
SI : Rata-rata indeks *Silhouette* dari dataset

SI_j : Rata-rata indeks *Silhouette* klaster ke-*j*

K : Jumlah klaster

C. Hasil Penelitian dan Pembahasan

Algoritma *K-Medoids* hadir untuk mengatasi masalah algoritma *K-Means* yang sensitif terhadap pencilan. Langkah awal dalam algoritma *K-Medoids* adalah menguji apakah dalam data terdapat pencilan atau tidak dengan menggunakan *boxplot*. Menurut Laraswati (2023), deteksi terhadap adanya *outlier* perlu dilakukan karena akan berdampak secara signifikan pada hasil analisis. Adapun *boxplot* untuk data penelitian adalah sebagai berikut:



Gambar 1. *Boxplot* Data Penelitian

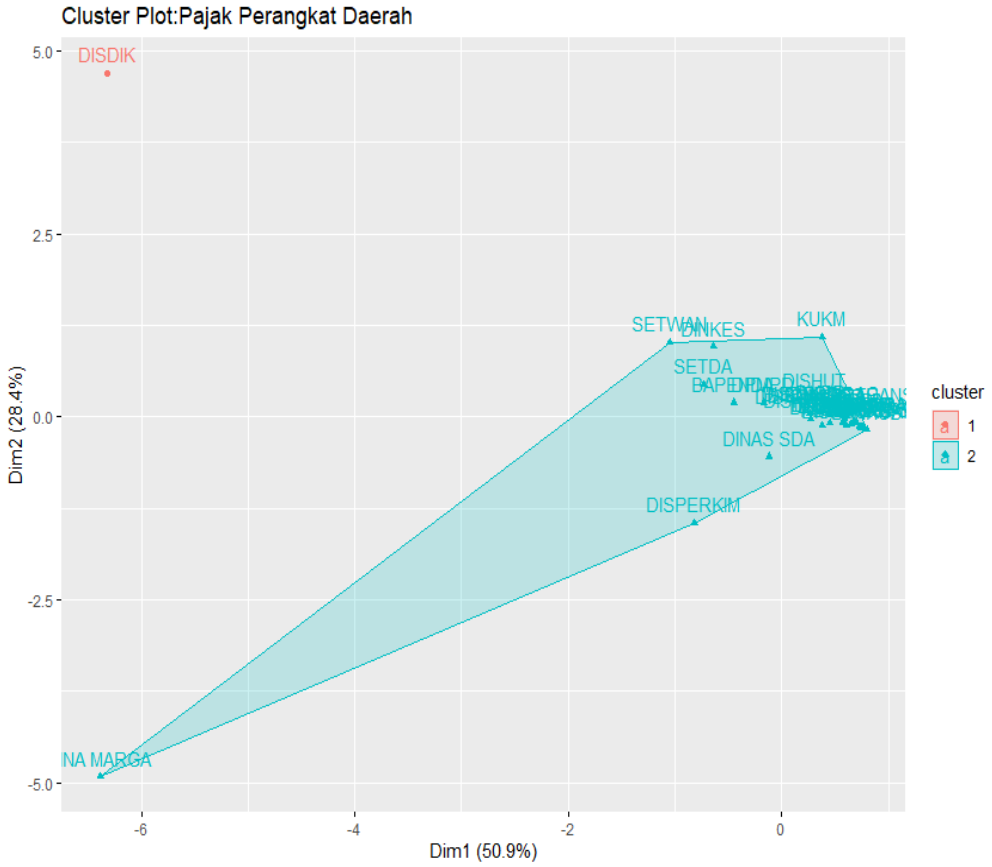
Berdasarkan Gambar 1 diatas, terlihat bahwa kelima variabel pada data memiliki pencilan dengan jumlah total sebanyak 27 data. Setelah dilakukan pengecekan *outlier*, langkah selanjutnya dalam algoritma *K-Medoids* adalah menentukan jumlah klaster yang akan digunakan. Pada penelitian ini penulis akan mencoba dengan menggunakan jumlah klaster K = 2, K = 3, K = 4, dan K = 5. Apabila jumlah klaster yang akan digunakan sudah ditentukan, langkah selanjutnya adalah melakukan analisis. Proses pengujian ini dilakukan dengan menggunakan *software* R-Studio dan diperoleh hasil sebagai berikut:

Skenario pengujian 1: *K-Medoids Clustering* dengan K = 2

Tabel 1. Rincian Anggota Klaster dengan K = 2

Klaster 1	Klaster 2
PD_1	Perangkat Daerah lainnya

Tabel 1 menunjukan hasil bahwa anggota klaster 1 sebanyak 1 perangkat daerah dan anggota klaster 2 sebanyak 37 perangkat daerah yang kemudian hasil tersebut di visulisasikan menjadi sebuah plot sebagaimana pada Gambar 2 berikut ini.



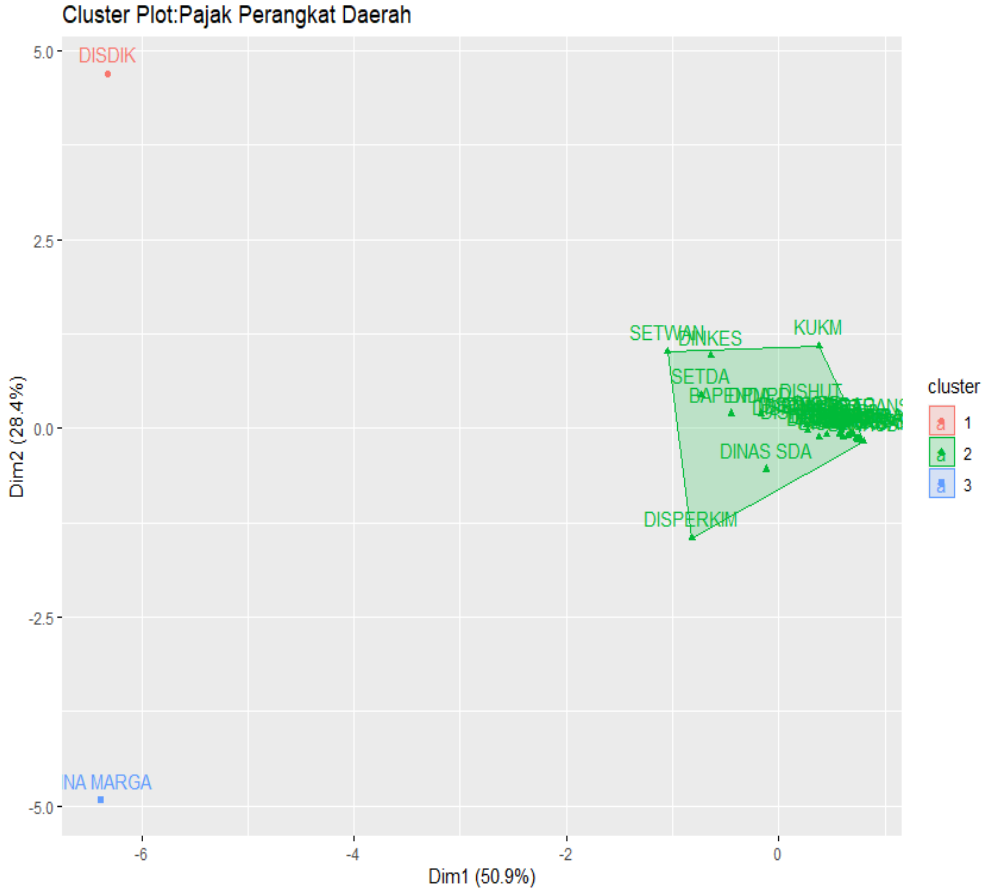
Gambar 2. Plot Anggota Klaster dengan K = 2

Skenario pengujian 2: *K-Medoids Clustering* dengan K = 3

Tabel 2. Rincian Anggota Klaster dengan K = 3

Klaster 1	Klaster 2	Klaster 3
PD_1	Perangkat Daerah Lainnya	PD_3

Tabel 2 menunjukan hasil bahwa anggota klaster 1 sebanyak 1 perangkat daerah, anggota klaster 2 sebanyak 36 perangkat daerah, dan anggota klaster 3 sebanyak 1 perangkat daerah yang kemudian hasil tersebut di visulisasikan menjadi sebuah plot sebagaimana pada Gambar 3 berikut ini.



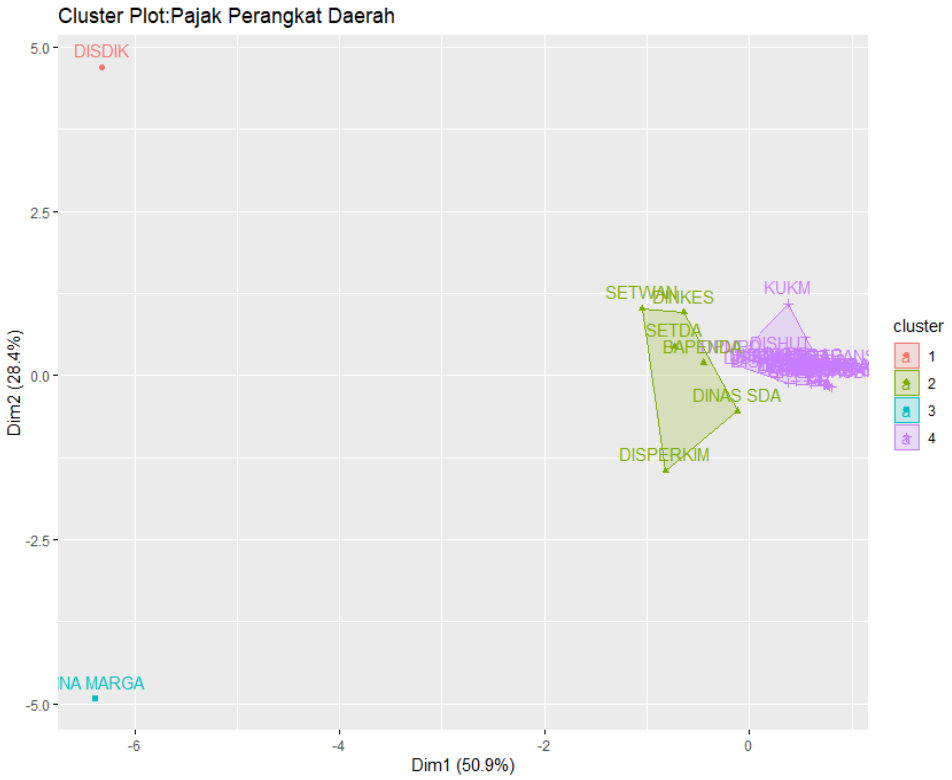
Gambar 3. Plot Anggota Klaster dengan K = 3

Skenario pengujian 3: *K-Medoids Clustering* dengan K = 4

Tabel 3. Rincian Anggota Klaster dengan K = 4

Klaster 1	Klaster 2	Klaster 3	Klaster 4
PD_1	PD_2, PD_4, PD_5, PD_{29}, PD_{30}, PD_{33}	PD_3	Perangkat Daerah Lainnya

Tabel 3 menunjukan hasil bahwa anggota klaster 1 sebanyak 1 perangkat daerah, anggota klaster 2 sebanyak 6 perangkat daerah, anggota klaster 3 sebanyak 1 perangkat daerah, dan anggota klaster 4 sebanyak 30 perangkat daerah yang kemudian hasil tersebut di visulisasikan menjadi sebuah plot sebagaimana pada Gambar 4 berikut ini.



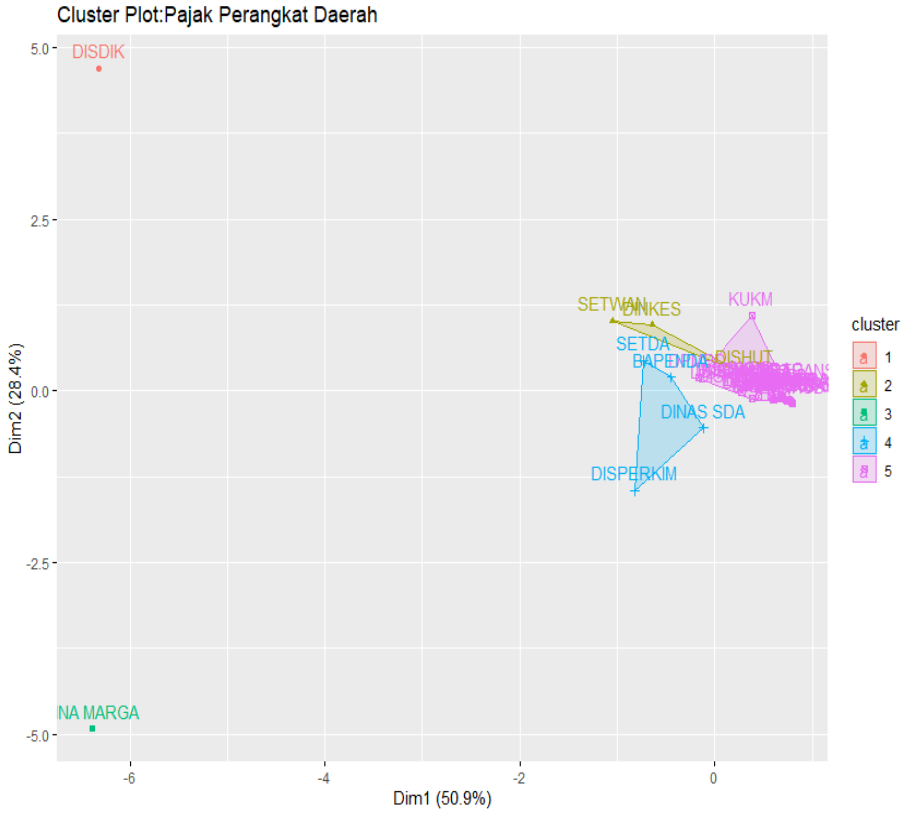
Gambar 4. Plot Anggota Klaster dengan K = 4

Skenario pengujian 4: *K-Medoids Clustering* dengan K = 5

Tabel 4. Rincian Anggota Klaster dengan K = 5

Klaster 1	Klaster 2	Klaster 3	Klaster 4	Klaster 5
PD_1	PD_2, PD_{26}, PD_{30}	PD_3	PD_4, PD_5, PD_{29}, PD_{33}	Perangkat Daerah Lainnya

Tabel 4 menunjukan hasil bahwa anggota klaster 1 sebanyak 1 perangkat daerah, anggota klaster 2 sebanyak 3 perangkat daerah, anggota klaster 3 sebanyak 1 perangkat daerah, anggota klaster 4 sebanyak 4 perangkat daerah, dan anggota klaster 5 sebanyak 29 perangkat daerah yang kemudian hasil tersebut di visulisasikan menjadi sebuah plot sebagaimana pada Gambar 5 berikut ini.



Gambar 5. Plot Anggota Klaster dengan K = 5

Selanjutnya dilakukan validasi klaster untuk melihat skenario pengujian mana yang paling baik. Karena analisis klaster hanya menunjukan anggota dari setiap klaster nya, bukan memutuskan jumlah klaster yang terbentuk. Adapun kriteria pengukuran tingkat kebaikan klaster menurut Faradilla (2022) yaitu pengambilan jumlah klaster dikatakan baik apabila memiliki Indeks *Connectivity* paling kecil, atau Indeks *Dunn* paling besar, dan/atau Indeks *Silhouette* paling besar. Berikut merupakan hasi dari validasi klaster nya:

Tabel 5. Hasil Validasi Klaster

Klaster	CI (*Connectivity Index)*	DI (*Dunn Index)*	SI *(Silhouette Index)*
2 Klaster	2,9290	1,0038	0,8700
3 Klaster	5,8579	2,0565	0,8591
4 Klaster	15,6234	0,1626	0,6293
5 Klaster	18,6357	0,2767	0,6325

Berdasarkan Tabel 5 diatas, maka dapat disimpulkan bahwa pembentukan 2 klaster merupakan yang terbaik. Sehingga penulis mengambil hasil pengklasteran dengan K = 2. Kemudian, untuk memahami karateristik masing-masing klaster dilakukan perhitungan rata-rata untuk setiap variabel dalam setiap klaster, dan diketahui bahwa klaster 1 memiliki rata-rata yang lebih tinggi dibandingkan dengan klaster 2. Oleh karena itu, dapat dikatakan bahwa anggota klaster 1 merupakan anggota klaster dengan tarikan pajak yang tinggi, sedangkan anggota klaster 2 merupakan anggota klaster dengan tarikan pajak yang rendah.

D. Kesimpulan

Berdasarkan hasil penelitian yang telah dilaksanakan, diperoleh kesimpulan bahwa hasil dari

proses pengelompokan perangkat daerah di Provinsi Jawa Barat berdasarkan data total rekapitulasi tarikan pajak di SIPD pada bulan Januari-Agustus tahun 2023 dengan menggunakan algoritma *K-Medoids Clustering* diperoleh sebanyak 2 klaster dengan jumlah anggota klaster 1 sebanyak 1 perangkat daerah yaitu PD_1 dan anggota klaster 2 sebanyak 37 perangkat daerah. Selain itu, diketahui juga nilai rata-rata tiap jenis pajak yang diberikan oleh klaster 1 lebih tinggi dibandingkan dengan nilai rata-rata tiap jenis pajak yang diberikan oleh klaster 2. Sehingga dapat dikatakan bahwa anggota klaster 1 merupakan anggota klaster dengan tarikan pajak yang tinggi, sedangkan anggota klaster 2 merupakan anggota klaster dengan tarikan pajak yang rendah. Hal ini dapat dijadikan acauan bagi instansi yang bersangkutan untuk lebih memperhatikan proses pengeluaran pajak agar dapat meminimalisir kesalahan dalam proses pencatatan transaksi dan sistem pelayanan dapat dikembangkan lagi sehingga lebih efisien, cepat, dan *modern*. Kemudian, untuk peneliti selanjutnya dapat mempelajari lagi metode lain dan mencari lebih banyak referensi agar dapat membandingkan hasilnya.

Ucapan Terimakasih

Puji dan syukur penulis panjatkan kehadirat Allah Subhanahu Wa Ta'ala atas limpahan rahmat dan karunia-Nya penulis dapat menyelesaikan jurnal artikel ini. Shalawat serta salam semoga selalu tercurah limpahkan kepada Nabi Muhammad SAW juga kepada keluarganya, para sahabatnya dan pengikutnya hingga akhir zaman. Penulis menyadari banyak pihak yang memberikan dukungan dan bantuan dalam menyelesaikan laporan ini. Oleh karena itu, sudah sepantasnya penulis mengucapkan terimakasih kepada Bapak Abdul Kudus, M.Si., Ph.D., selaku Ketua Prodi Statistika FMIPA Universitas Islam Bandung dan dosen Pembimbing dalam pembuatan jurnal artikel ini. Dengan adanya laporan ini diharapkan dapat bermanfaat bagi para pembaca dan dapat dijadikan referensi demi pengembangan ke arah yang lebih baik. Dalam penulisan laporan ini penulis menyadari masih terdapat kekurangan, penulis memohon maaf dan mengharapkan adanya kritik dan saran sebagai bahan evaluasi untuk penyempurnaan laporan ini.

Daftar Pustaka

Nurfauzan, A. F., & Darwis, S. (2024). Clustering Probabilistic Seismic Hazard Analysis Temporal Epidemic-Type Aftershock Sequence untuk Premi Asuransi. *Jurnal Riset Statistika*, 41–48. https://doi.org/10.29313/jrs.v4i1.3864

Afira, Natasya., & Wijayanto, Arie W. (October, 2021). *Analisis Cluster Kemiskinan Provinsi di Indonesia Tahun 2019 dengan Metode Partitioning dan Hierarki.* Komputika: Jurnal Sistem Komputer, Vol. 10, No. 2, 101-109. 10.34010/komputika.v10i2.4317.

Faradilla, Shafa Bunga. (2022). *Komparasi Analisis K-Medoids Clustering dan Hierarchical Clustering*. Tugas Akhir (S1), Universitas Islam Indonesia Yogyakarta.

Laraswati, B. D. (Maret 8, 2023). *5 Cara Mendeteksi Outlier dalam Data*. Algoritma Data Science School. Retrieved October 4, 2023, from https://blog.algorit.ma/cara-mendeteksi-outlier/.

Nurkholik, David. (2022). *Analisis K-Medoids Clustering Metode Elbow pada Kasus Covid-19 di Provinsi DKI Jakarta.* Skripsi, Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Paramadina, M., Sudarmin, & Aidid, M. K. (2019). *Perbandingan Analisis Cluster Metode Average Linkage dan Metode Ward (Kasus: IPM Provinsi Sulawesi Selatan)*. *V*ARIANSI: Journal of Statistics and Its Application on Teaching and Research, Vol. 1, No. 2, 23-31. 10.35580/variansiunm9357

Sukmawati (2017). *Analisis Cluster dengan Metode Hierarki untuk Pengelompokkan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Indikator Makro Ekonomi.* Skripsi tidak dipublikasikan. Program Sarjana, Program Studi Matematika, Universitas Islam Negeri Alauddin Makassar.

Widyadhana, D., dkk. (2021). *Perbandingan Analisis Klaster K-Means dan Average Linkage untuk Pengklasteran Kemiskinan di Provinsi Jawa Tengah*. PRISMA, Prosiding Seminar Nasional Matematika, 4, 584-594.

Zulfa, Farida. (2019). *Pengelompokan Provinsi Di Indonesia Berdasarkan Indikator Pendidikan Menggunakan K-Means Dan K-Medoids.* Sarjana Terapan (S1/D4) tesis, Universitas Negeri Semarang.