PERBANDINGAN METODE K-MEANS DAN K- MEDOIDS DALAM CLUSTERING FAKTUR PAJAK DI PT AGUNG SERVIS JAYA SAKTI

Nanda Tamala Sari, Talitha Desya Salsabila, Aliya Rafida Zahra, Jefina Tri Kumalasari
Sistem Informasi, Universitas Bina Sarana Informatika
Jl. Kramat Raya No.98 Senen, Jakarta Pusat
triascendekia@gmail.com

ABSTRAK

PT Agung Servis Jaya Sakti merupakan perusahaan yang bergerak di bidang penyewaan dan penjualan *forklift* dengan volume transaksi yang tercatat dalam bentuk faktur pajak elektronik (*e-Faktur*) melalui sistem *CORETAX*. Besarnya jumlah data transaksi menyebabkan perusahaan mengalami kesulitan dalam mengidentifikasi pola pelanggan secara manual. Penelitian ini bertujuan untuk mengelompokkan data faktur pajak pelanggan menggunakan metode K-Means dan K-Medoids serta membandingkan kualitas hasil *clustering* yang dihasilkan oleh kedua metode tersebut. Variabel yang digunakan meliputi Nilai Ekonomi Transaksi (DPP), Besaran Beban Pajak (PPN), dan Periode Masa Pajak. Proses penelitian meliputi tahap seleksi data, pembersihan data, transformasi, normalisasi, proses *clustering* menggunakan K-Means dan K-Medoids, serta evaluasi kualitas *cluster* menggunakan *Davies–Bouldin Index* (DBI). Hasil pengujian menunjukkan bahwa metode K-Means menghasilkan nilai DBI sebesar 0,380, sedangkan metode K-Medoids menghasilkan nilai DBI sebesar 1,141. Nilai DBI yang lebih rendah menunjukkan kualitas *cluster* yang lebih baik, sehingga K-Means menghasilkan pemisahan *cluster* yang lebih optimal dibandingkan K-Medoids. Namun demikian, K-Medoids menunjukkan kestabilan *cluster* yang lebih baik terhadap keberadaan data ekstrem. Hasil penelitian ini menunjukkan bahwa pengelompokan data faktur pajak mampu membentuk tiga kelompok utama pelanggan, yaitu pelanggan dengan transaksi rendah, sedang, dan tinggi, yang dapat digunakan sebagai dasar pendukung pengambilan keputusan perpajakan dan strategi bisnis perusahaan.

Kata kunci : *Clustering, K-Means, K-Medoids, Faktur Pajak, Rapid Miner, Davies–Bouldin Index*

1. PENDAHULUAN

Perkembangan teknologi informasi telah mendorong digitalisasi administrasi perpajakan untuk meningkatkan efisiensi, akurasi, dan transparansi dalam proses pencatatan serta pelaporan pajak. Salah satu implementasinya adalah sistem *CORETAX*, yaitu sistem administrasi perpajakan berbasis data yang memungkinkan perusahaan mengelola dan memantau faktur pajak secara elektronik. Meskipun sistem ini menghasilkan data dalam jumlah besar secara terstruktur, pemanfaatannya masih memerlukan metode analisis yang tepat agar dapat memberikan nilai tambah bagi perusahaan.

PT Agung Servis Jaya Sakti sebagai perusahaan yang bergerak di bidang penyewaan dan penjualan *forklift* menghasilkan data faktur pajak dengan volume besar dan variabel utama seperti DPP, PPN, dan masa pajak. Besarnya volume dan kompleksitas data menyebabkan analisis manual menjadi kurang efektif dan berpotensi menimbulkan kesalahan, sehingga perusahaan mengalami kesulitan dalam mengidentifikasi pola transaksi pelanggan.

Kondisi tersebut mengakibatkan data faktur pajak belum dimanfaatkan secara optimal sebagai dasar pengambilan keputusan. Tanpa adanya pengelompokan data yang sistematis, perusahaan menghadapi kendala dalam menentukan prioritas pelanggan, mengidentifikasi pelanggan dengan kontribusi transaksi terbesar, serta mendeteksi pola transaksi yang tidak wajar.

Untuk mengatasi permasalahan tersebut, digunakan pendekatan *data mining*, yaitu proses penggalian informasi dan pola penting dari kumpulan data menggunakan metode statistik, matematika, dan kecerdasan buatan [1]. Salah satu teknik yang relevan adalah *clustering*, yaitu proses pengelompokan data berdasarkan tingkat kemiripan karakteristik.

Dua metode *clustering* yang umum digunakan adalah K-Means dan K-Medoids. K-Means memiliki keunggulan dalam kecepatan *komputasi* tetapi sensitif terhadap *outlier* karena pusat *cluster* ditentukan berdasarkan nilai rata-rata, sedangkan K-Medoids lebih stabil terhadap *outlier* karena menggunakan objek aktual sebagai pusat *cluster* [2].

Berdasarkan studi literatur, penelitian yang secara khusus membandingkan K-Means dan K-Medoids pada data faktur pajak elektronik berbasis sistem *CORETAX* masih terbatas. Sebagian besar penelitian sebelumnya menerapkan *clustering* pada data penjualan, kinerja pegawai, dan faktor sosial lainnya [3]–[8], sehingga terdapat *research gap* dalam penerapan teknik *clustering* pada *domain* perpajakan digital.

Oleh karena itu, penelitian ini bertujuan untuk menerapkan dan membandingkan metode K-Means dan K-Medoids dalam pengelompokan data faktur pajak berbasis sistem *CORETAX* di PT Agung Servis Jaya Sakti serta mengevaluasi kualitas cluster menggunakan *Davies–Bouldin Index* (DBI). Penelitian dilakukan melalui tahapan pengumpulan data, data

cleaning, transformasi, normalisasi, proses *clustering*, dan evaluasi DBI.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Beberapa penelitian sebelumnya telah menerapkan metode *clustering* menggunakan algoritma K-Means, K-Medoids, maupun kombinasi keduanya serta menggunakan *Davies–Bouldin Index* (DBI) sebagai metrik evaluasi. Penelitian - penelitian tersebut menjadi dasar metodologis bagi penelitian ini meskipun sebagian besar dilakukan pada domain yang berbeda dari perpajakan digital.

a. Pada penelitian ini menerapkan algoritma K-Means pada data penilaian kinerja 455 karyawan dengan pengujian K = 2 dan memperoleh nilai DBI sebesar –1,752.[1]
b. Pada penelitian ini membandingkan kinerja algoritma X-Means, K-Means, dan K-Medoids dalam pengelompokan data pegawai.[2]
c. Pada penelitian ini menggunakan K-Means untuk menganalisis faktor penyebab perceraian dan memperoleh hasil terbaik pada K = 7 dengan nilai DBI sebesar –0,337.[3]
d. Pada penelitian ini membandingkan performa K-Means dan K-Medoids dalam pengelompokan data produksi susu segar di Indonesia.[4]
e. Pada penelitian ini menerapkan K-Means dan K-Medoids pada data penjualan untuk menentukan produk unggulan dan memperoleh nilai DBI masing-masing sebesar –0,430 dan –1,392.[5]
f. Pada penilitian ini melakukan analisis data *Netflix* menggunakan *RapidMiner* yang mencakup tahap *preprocessing*, pemodelan, dan *eksplorasi* pola pengguna.[6]
g. Sementara itu, penelitian ini menerapkan K-Means pada data transaksi penjualan pakaian tidak terstruktur dan berhasil membentuk beberapa *cluster* berdasarkan frekuensi pembelian dan nilai transaksi.[7]

2.2. *Data Mining*

Data mining digunakan untuk menggali pola dan informasi penting dari kumpulan data sesuai tahapan *Knowledge Discovery in Database* (KDD) [8]. Dalam penelitian ini, *data mining* dimanfaatkan untuk mengolah data faktur pajak berbasis sistem *CORETAX* guna mengidentifikasi pola transaksi pelanggan yang dapat digunakan sebagai dasar pengambilan keputusan perusahaan.

2.3. *Clustering*

Clustering merupakan teknik dalam data mining untuk mengelompokkan data berdasarkan tingkat kemiripan karakteristik tanpa memerlukan label awal [9]. Pada penelitian ini, *clustering* diterapkan pada data faktur pajak *CORETAX* berdasarkan nilai DPP, PPN, dan periode masa pajak untuk memperoleh kelompok pelanggan dengan pola transaksi yang berbeda.

2.4. *Coretax*

Coretax adalah sistem inti administrasi perpajakan yang memiliki kemampuan untuk terintegrasi dengan sistem akuntansi yang digunakan oleh perusahaan maupun individu.[10]

Gambar 1. *Coretax*

Data dari sistem ini digunakan sebagai sumber utama dalam penelitian, karena merepresentasikan aktivitas transaksi dan pelaporan pajak perusahaan secara digital dan terstruktur.

2.5. *RapidMiner*

RapidMiner digunakan sebagai alat bantu untuk melakukan pengolahan data, penerapan algoritma K-Means dan K-Medoids, serta evaluasi hasil *clustering* menggunakan *Davies–Bouldin Index* (DBI) secara terintegrasi.[11]

2.6. Metode K-Means

Metode K-Means mengelompokkan data berdasarkan kedekatan terhadap pusat *cluster* yang dihitung menggunakan nilai rata-rata [12]. Dalam penelitian ini, K-Means digunakan untuk mengelompokkan data faktur pajak berdasarkan nilai transaksi dan pajak guna membentuk kelompok pelanggan dengan karakteristik tertentu.

2.7. Metode K-Medoids

Metode K-Medoids menggunakan objek aktual sebagai pusat *cluster* sehingga lebih stabil terhadap data ekstrem dibandingkan K-Means [13]. Metode ini digunakan sebagai pembanding untuk memperoleh hasil *clustering* yang lebih representatif pada data faktur pajak.

2.8. *Davies Bouldien Index* (DBI)

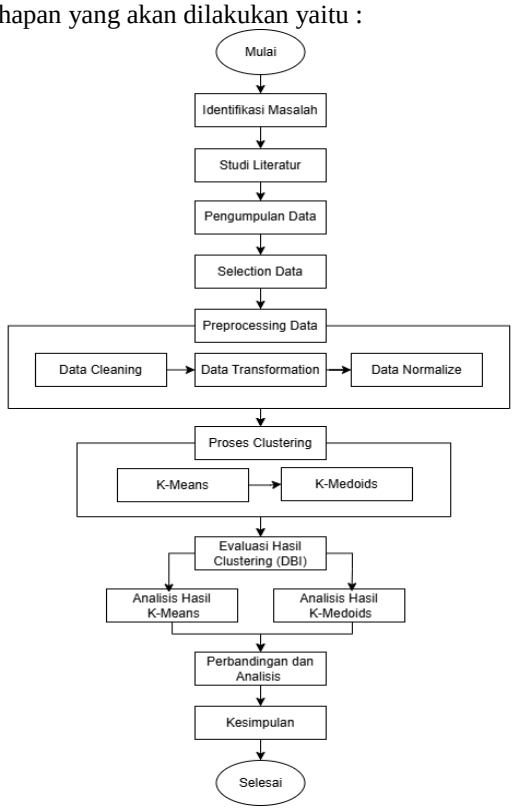
Davies–Bouldin Index (DBI) digunakan untuk mengevaluasi kualitas hasil *clustering* berdasarkan rasio kedekatan *intra-cluster* dan jarak antar-*cluster*. Nilai DBI yang lebih kecil menunjukkan kualitas *cluster* yang lebih baik, sehingga digunakan untuk membandingkan hasil *clustering* antara metode K-Means dan K-Medoids.

3. METODE PENELITIAN

Penyusunan kerangka penelitian ini dibuat dengan mengacu pada permasalahan yaitu belum adanya analisis pengelompokan pelanggan berdasarkan data faktur pajak digital *CORETAX*, serta perlunya perbandingan kinerja algoritma K-Means dan K-Medoids pada konteks data perpajakan yang bersifat *variatif* dan berpotensi mengandung *outlier*.

Kerangka ini mencakup tahapan dari awal hingga akhir penelitian, dengan fokus pada pengolahan data faktur pajak menggunakan dua metode *clustering* (K-Means dan K-Medoids).

Tahapan divisualisasikan dalam diagram alir untuk memudahkan pemahaman, dan meliputi semua langkah yang akan dilakukan secara menyeluruh. Tahapan yang akan dilakukan yaitu :



Gambar 2. Kerangka Penelitian

Gambat di atas menunjukkan alur proses penelitian yang terdiri beberapa tahapan yaitu tahap pengumpulan data, *preprocessing*, *clustering*, evaluasi, analisis dan kesimpulan.

3.1. Identifikasi Masalah

Proses penelitian diawali dengan tahapan identifikasi masalah guna memperoleh gambaran awal terkait kondisi yang diteliti di PT Agung Servis Jaya Sakti, yaitu belum adanya analisis pengelompokan pelanggan berdasarkan data faktur pajak yang dapat membantu dalam pengambilan keputusan strategis

3.2. Studi *Literatur*

Peneliti menelaah *literatur* terkait *data mining* dan teknik *clustering* sebagai dasar metodologis penelitian. *Data mining* dipahami sebagai proses sistematis untuk menemukan pola tersembunyi dalam data berukuran besar

3.3. Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa data sekunder yang di peroleh dari sistem *CORETAX* pada PT Agung Servis Jaya Sakti

3.4. *Selection* Data

Tahap pemilihan data dilakukan untuk memastikan hanya variabel yang relevan dengan analisis penelitian yang digunakan.

3.5. *Preprocessing* Data

Tahap *preprocessing* dilakukan untuk memastikan bahwa data yang digunakan dalam proses *clustering* berada dalam kondisi bersih, konsisten, dan siap diproses oleh algoritma. Tahap ini meliputi data *cleaning*, data *transformation*, dan normalisasi, sehingga hasil pengelompokan dapat lebih akurat.

Normalize (*Min–Max Scaling*) Normalisasi diperlukan untuk menyamakan skala variable.Proses normalisasi dilakukan menggunakan metode *Min–Max Scaling, d*engan rumus sebagai berikut*:*

$$X' = \frac{X - Xmin}{Xmax - Xmin} \quad (1)$$

3.6. Proses *Clustering*

Pada tahap ini dilakukan proses pengelompokan (*clustering*) menggunakan dua algoritma yaitu K-Means dan K-Medoids. K-Means merupakan algoritma *clustering* berbasis *centroid* yang menggunakan nilai rata-rata (*mean*) sebagai pusat *cluster*, sehingga cenderung lebih sensitif terhadap *outlier*. Sebaliknya, K-Medoids menggunakan objek aktual (*medoid*) sebagai pusat *cluster* sehingga lebih stabil pada dataset dengan variasi nilai transaksi yang besar.

Kedua algoritma diterapkan untuk membentuk tiga *cluster* pelanggan berdasarkan pola transaksi faktur pajak. Penetapan jumlah *cluster* K = 3 tidak ditentukan oleh perhitungan DBI, tetapi didasarkan pada tujuan segmentasi pelanggan. Segmentasi ini mencakup:

a. *Cluster* C0 : transaksi dan pajak rendah, awal masa pajak
b. *Cluster* C1 : transaksi dan pajak sedang, pertengahan masa pajak
c. *Cluster* C2 : transaksi dan pajak tinggi, akhir masa pajak

Penentuan ini bersifat *konseptual* dan selanjutnya divalidasi secara kuantitatif menggunakan *Davies–Bouldin Index* (DBI) guna memastikan kualitas pengelompokan yang optimal.

3.7. Metode K-Means

Algoritma K-Means merupakan salah satu algoritma clustering yang bekerja dengan cara mengelompokkan data berdasarkan titik pusat *cluster* atau *centroid*

Menurut beberapa penelitian, langkah-langkah penerapan algoritma K-Means umumnya meliputi [12], [14]:

a. Tentukan jumlah *cluster* k (pada penelitian ini *k* = 3 berdasarkan tujuan segmentasi).
b. Menentukan *centroid* awal secara acak dari data yang tersedia sebanyak jumlah *cluster* (k).
c. Menghitung jarak antara titik *centroid* dengan titik tiap objek :

$$D(X,Y) = \sqrt{\sum_{i=1}^{n} (xi - yi)^2} \quad (2)$$

Keterangan :
D(X,Y) = Jarak *Euclidean* antara titik data X dan *medoid* Y.
X = Titik data (objek) pelanggan yang diukur jaraknya.
Y = *Medoid* (pusat *cluster* berupa data aktual).
n = Jumlah variabel (n = 3: DPP, PPN, Masa Pajak).

d. Lakukan *iterasi*, kemudian tentukan *centeroid* baru.
e. Ulangi sampai *centeroid* yang terbentuk tidak berubah lagi.

3.8. Metode K-Medoids

K-Medoids merupakan data asli yang dijadikan perwakilan dari suatu *cluster*

Pada setiap *cluster*, K-Medoids merupakan data yang memiliki jarak total paling kecil terhadap seluruh data lainnya di dalam *cluster*. Sebelum menghitung jarak antar data dengan parameter berbeda, sebaiknya dilakukan normalisasi agar skala data seimbang dan hasil *clustering* lebih akurat

Berbeda dengan K-Means, K-Medoids bukan merupakan titik pusat geometris dari *cluster* tersebut. berikut langkah-langkah K-Medoids

a. *Inisialisasi* dimulai dengan menetapkan K titik data secara acak dari dataset yang akan berperan sebagai *medoid* awal.
b. Pada tahap *assignment*, setiap data diklasifikasikan ke *medoid* terdekat, di mana jarak antar data dan *medoid* diukur menggunakan metrik *Euclidean Distance*. Pada tahap pembaharuan, total jarak setiap *medoid* terhadap seluruh titik data dalam klaster dihitung, dan *medoid* baru ditetapkan sebagai titik data yang memiliki total jarak minimum.
c. Tahap *iterasi* melibatkan pengulangan penugasan dan pembaharuan *medoid* sampai tidak ada perubahan lagi.
d. Setelah itu, partisi klaster akhir beserta *medoid* dan titik data yang tergabung ditetapkan sebagai output.

3.9. Evaluasi Hasil *Clustring*

Evaluasi hasil *clustering* dilakukan menggunakan *Davies–Bouldin Index* (DBI) untuk menilai kualitas pemisahan antar *cluster*.
Perhitungan DBI menggunakan rumus sebagai berikut:

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \frac{max}{j\neq 1}\left(\frac{Si+Sj}{Mij}\right) \quad (3)$$

3.10. Analisis Hasil K-Means

Analisis hasil K-Means dilakukan untuk mengetahui pola pengelompokan data faktur pajak berdasarkan nilai DPP, PPN, dan Masa Pajak. Hasil *clustering* dianalisis dengan melihat distribusi data pada setiap *cluster*, karakteristik masing-masing *cluster*, serta kualitas pengelompokan yang diukur menggunakan *Davies–Bouldin Index* (DBI). Nilai DBI digunakan sebagai indikator untuk menilai tingkat kekompakan dan pemisahan antar *cluster* yang dihasilkan oleh metode K-Means.

3.11. Analisis Hasil K-Medoids

Analisis hasil K-Medoids dilakukan untuk mengevaluasi pola pengelompokan data faktur pajak dengan menggunakan objek aktual sebagai pusat *cluster* (*medoid*). Analisis difokuskan pada distribusi data dalam setiap *cluster*, stabilitas hasil pengelompokan terhadap keberadaan *outlier*, serta kualitas *cluster* yang dihasilkan berdasarkan nilai *Davies–Bouldin Index* (DBI). Hasil analisis ini digunakan untuk membandingkan performa metode K-Medoids dengan metode K-Means.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Penelitian ini menggunakan dataset berupa 248 data faktur pajak elektronik (*e-Faktur*) yang bersumber dari sistem *CORETAX* milik PT Agung Jaya Sakti untuk periode Januari hingga September 2025. Data disajikan dalam format Microsoft Excel dan terdiri atas 17 atribut transaksi perpajakan, yang selanjutnya digunakan sebagai dasar dalam proses pemodelan *clustering* dengan metode K-Means dan K-Medoids.

Proses *clustering* difokuskan pada tiga atribut utama, yaitu Nilai Ekonomi Pajak (DPP), Besaran Beban Pajak (PPN), dan Periode Masa Pajak. DPP merepresentasikan nilai ekonomi transaksi, PPN mencerminkan besaran pajak yang dikenakan, sedangkan Masa Pajak berfungsi sebagai dimensi waktu untuk mengidentifikasi pola transaksi. Kombinasi ketiga atribut tersebut memungkinkan pengelompokan faktur pajak dilakukan secara *representatif* berdasarkan kemiripan nilai transaksi, beban pajak, dan periode pelaporan.

4.2. *Selection* Data

Tahap data *selection* dilakukan dengan memilih atribut yang relevan dari data mentah faktur pajak untuk mendukung proses *clustering*. Atribut yang digunakan meliputi Nama Pembeli, Masa Pajak, DPP Nilai Lain/DPP, dan PPN, di mana Nama Pembeli berfungsi sebagai identitas data, sedangkan Masa Pajak, DPP, dan PPN merepresentasikan dimensi waktu, nilai transaksi, dan besaran pajak. Seleksi atribut ini bertujuan menyederhanakan dataset agar lebih terfokus dan siap digunakan pada tahap data *cleaning*, transformasi, dan normalisasi.

4.3. *Preprocessing* Data

Tahap *preprocessing* data meliputi proses data *cleaning*, transformasi, dan normalisasi.

a. Data *Cleaning*
Pada tahap data *cleaning*, dilakukan pemeriksaan terhadap atribut Nama Pembeli, DPP Nilai Lain/DPP, PPN, dan Masa Pajak untuk memastikan tidak terdapat nilai kosong (*missing values*). Data yang tidak lengkap dihapus sehingga

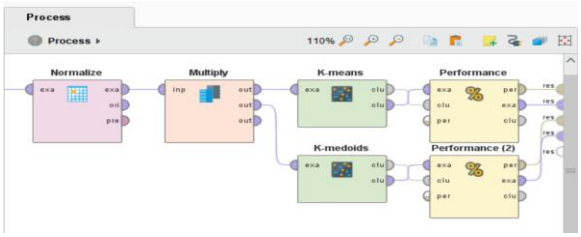
dataset yang digunakan bersifat bersih dan konsisten, guna menghindari kesalahan perhitungan dalam analisis *clustering*.

b. Transformasi
Tahap transformasi data bertujuan mengubah format data agar dapat diproses secara numerik. Atribut DPP dan PPN yang semula mengandung simbol mata uang dan pemisah angka diubah menjadi nilai numerik murni, sedangkan atribut Masa Pajak dikonversi ke dalam bentuk numerik untuk merepresentasikan periode waktu

c. *Normalize*.
Selanjutnya, dilakukan normalisasi data menggunakan metode *Min*–Max *Normalization* pada variabel DPP, PPN, dan Masa Pajak untuk menyamakan skala nilai dalam rentang 0 hingga 1. Proses normalisasi ini memastikan setiap variabel memiliki kontribusi yang seimbang dalam perhitungan jarak pada metode K-Means dan K-Medoids, sehingga data hasil *preprocessing* siap digunakan pada tahap clustering dan evaluasi kualitas *cluster* menggunakan *Davies–Bouldin Index* (DBI).

4.4. Proses Clustering

Gambar berikut menunjukkan model proses *clustering* K-Means dan K-Medoids yang digunakan dalam penelitian ini.



Gambar 3. Model *Clustering* K-Means dan K-Medoids

Dengan k = 3 dan *random seed* 54, kedua metode menghasilkan tiga cluster yang konsisten, dengan evaluasi kualitas *cluster* berdasarkan nilai *Davies–Bouldin Index* (DBI).

4.5. Evaluasi Hasil *Clustering* DBI

a. Evaluasi DBI jumlah *cluster* K (K2-K10)

Tabel 1. Evaluasi DBI K (K2-K10)

K	DBI K-Means	BDI K-Medoids
K2	0,877	1,657
K3	0,380	1,141
K4	0,413	0,813
K5	0,464	1,652
K6	0,525	1,199
K7	0,557	0,847
K8	0,491	0,926
K9	0,469	0,985
K10	0,517	1,090

Berdasarkan Tabel 1, evaluasi kualitas *clustering* menggunakan *Davies–Bouldin Index* (DBI) menunjukkan bahwa nilai DBI terendah pada metode K-Means diperoleh pada K = 3 sebesar (0,380), setelah mengalami penurunan dari K = 2 (0,877), sedangkan pada jumlah *cluster* di atas K = 3 nilai DBI kembali meningkat. Pada metode K-Medoids, nilai DBI terendah juga diperoleh pada K = 3, yaitu sebesar 1,141, dan tidak terdapat nilai DBI yang lebih rendah pada variasi jumlah *cluster* lainnya. Dengan demikian, jumlah *cluster* optimal untuk kedua metode adalah K = 3 karena menghasilkan kualitas *clustering* terbaik yang ditunjukkan oleh nilai DBI terendah

b. Evaluasi DBI jumlah cluster K3 K-Means

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -0.025
Avg. within centroid distance_cluster_0: -0.031
Avg. within centroid distance_cluster_1: -0.020
Avg. within centroid distance_cluster_2: -0.015
Davies Bouldin: -0.380
```

Gambar 4. PerformanceVector K-Means (K3)

Pada Gambar 4 hasil evaluasi menunjukkan bahwa nilai *Davies–Bouldin Index* (DBI) sebesar 0,380, yang menandakan bahwa hasil *clustering* memiliki kualitas yang baik. Nilai DBI yang semakin rendah menunjukkan bahwa cluster yang terbentuk memiliki tingkat kekompakan internal yang tinggi serta pemisahan antar *cluster* yang cukup jelas. Selain itu, nilai *average within centroid distance* sebesar 0,025 mengindikasikan bahwa jarak rata-rata data terhadap *centroid* relatif kecil, sehingga data dalam masing-masing *cluster* cenderung berdekatan dengan pusat *clusternya*.

Secara rinci, nilai *average within centroid distance* pada setiap *cluster* menunjukkan bahwa *cluster* 0 memiliki nilai 0,031, *cluster* 1 sebesar 0,020, dan cluster 2 sebesar 0,015. Nilai ini menggambarkan bahwa *cluster* 0 memiliki tingkat kekompakan paling tinggi dibandingkan *cluster* lainnya, sedangkan *cluster* 2 memiliki tingkat kekompakan yang relatif lebih rendah.

c. Evaluasi DBI jumlah *cluster* K3 K-Medoids

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -0.062
Avg. within centroid distance_cluster_0: -0.013
Avg. within centroid distance_cluster_1: -0.005
Avg. within centroid distance_cluster_2: -0.098
Davies Bouldin: -1.141
```

Gambar 5. PerformanceVector K-Medoids (K3)

Berdasarkan Gambar 5 *Nilai average within centroid distance* sebesar 0,062 menunjukkan bahwa secara keseluruhan jarak rata-rata data terhadap medoid relatif kecil, sehingga data dalam setiap *cluster* cenderung berdekatan dengan pusat *clusternya*. Secara rinci, nilai *average* within *centroid distance* pada masing-masing *cluster* adalah 0,013 untuk *cluster* 0, 0,005 untuk *cluster* 1, dan 0,098 untuk cluster 2. Nilai ini menunjukkan bahwa *cluster* 2 memiliki tingkat kekompakan paling tinggi, sedangkan *cluster* 1 memiliki tingkat kekompakan yang relatif lebih rendah dibandingkan *cluster* lainnya.

4.6. Hasil dan Analisis *Clustering* K-Means (K=3)

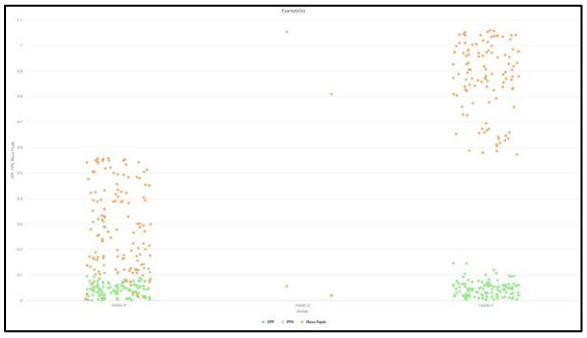
Tabel 2. Hasil akhir Clustering K-means

Cluster	DPP	PPN	Masa Pajak
C0	0,032	0,032	0,281
C1	0,024	0,024	0,869
C2	0,915	0,915	0,000

Pada Tabel 2 *Cluster* C0 memiliki nilai *centroid* DPP sebesar 0,032, PPN sebesar 0,032, dan Masa Pajak sebesar 0,281. *Cluster* ini menggambarkan kelompok transaksi dengan nilai ekonomi dan beban pajak yang relatif rendah yang cenderung terjadi pada periode awal masa pajak, sehingga didominasi oleh transaksi berskala kecil dengan kontribusi pajak yang rendah.

Cluster C1 memiliki nilai *centroid* DPP sebesar 0,024, PPN sebesar 0,024, dan Masa Pajak sebesar 0,869. Nilai Masa Pajak yang tinggi menunjukkan bahwa transaksi dalam *cluster* ini lebih banyak terjadi pada periode akhir masa pajak, meskipun nilai DPP dan PPN masih tergolong rendah hingga menengah.

Sementara itu, *cluster* C2 memiliki nilai *centroid* DPP sebesar 0,915, PPN sebesar 0,915, dan Masa Pajak sebesar 0,000. *Cluster* ini merepresentasikan transaksi dengan nilai DPP dan PPN yang sangat tinggi, yang menunjukkan transaksi berskala besar dengan beban pajak yang signifikan dan cenderung terjadi pada awal periode pelaporan pajak.



Gambar 6. *Scatter Plot* hasil *clustering* K-Means

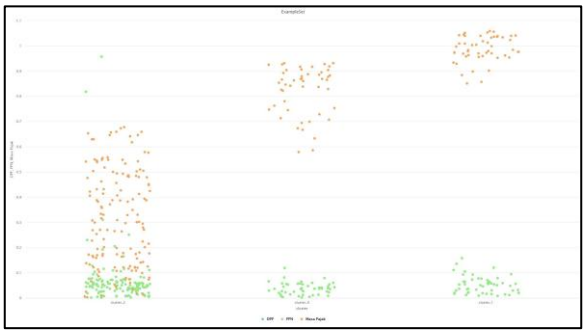
Pada Gambar 6 Grafik *scatter plot* digunakan untuk memvisualisasikan persebaran data dan pemisahan antar *cluster*. Setiap warna pada grafik merepresentasikan satu *cluster* yang berbeda. Jarak antar titik menunjukkan tingkat kemiripan data, di mana titik yang berdekatan memiliki karakteristik transaksi yang serupa. Pemisahan warna yang semakin jelas menunjukkan kualitas cluster yang semakin baik. Menunjukkan *scatter plot* hasil *clustering* K-Means yang memperlihatkan bahwa *cluster* 0 dan *cluster* 1 memiliki distribusi atribut yang relatif berdekatan pada variabel PPN (nilai rendah), sementara *cluster* 2 hanya terdiri dari dua data dengan nilai ekstrem sehingga dianggap sebagai *outlier*. Pola ini menunjukkan bahwa K-Means cenderung sensitif terhadap *outlier*, mengakibatkan satu *cluster* berisi sangat sedikit data

4.7. Hasil dan Analisis *Clustering* K-Medoids (K=3)

Tabel 3. Hasil akhir *Clustering* K-medoids

Cluster	DPP	PPN	Masa Pajak
C0	0,001	0,001	0,750
C1	0,054	0,054	1,000
C2	0,018	0,018	0,000

Pada Tabel 3 Berdasarkan nilai *medoid* akhir, *cluster* C0 memiliki nilai DPP dan PPN masing-masing sebesar 0,001 serta Masa Pajak sebesar 0,750, yang menunjukkan transaksi dengan nilai ekonomi dan beban pajak sangat rendah yang tersebar pada pertengahan hingga akhir masa pajak. *Cluster* C1 memiliki nilai *medoid* DPP dan PPN tertinggi, masing-masing sebesar 0,054, dengan nilai Masa Pajak sebesar 1,000, sehingga merepresentasikan transaksi bernilai tinggi yang dominan terjadi pada akhir masa pajak. Sementara itu, *cluster* C2 memiliki nilai *medoid* DPP dan PPN sebesar 0,018 serta Masa Pajak sebesar 0,500, yang menunjukkan transaksi dengan nilai ekonomi dan beban pajak pada tingkat menengah yang tersebar pada periode pertengahan masa pajak. Secara keseluruhan, metode K-Medoids mampu mengelompokkan data faktur pajak secara stabil berdasarkan nilai transaksi, beban pajak, dan periode masa pajak karena pusat *cluste*r ditentukan oleh data aktual (*medoid*).



Gambar 7. *Scatter Plot* hasil *clustering* k-medoids

Berdasarkan Gambar 7 Grafik *scatter plot* digunakan untuk memvisualisasikan persebaran data dan pemisahan antar *cluster*. Setiap warna pada grafik merepresentasikan satu *cluster* yang berbeda. Jarak antar titik menunjukkan tingkat kemiripan data, di mana titik yang berdekatan memiliki karakteristik transaksi yang serupa. Pemisahan warna yang semakin jelas menunjukkan kualitas *cluster* yang semakin baik.

Pada gambar grafik *scatter plot* hasil K-Medoids, distribusi data pada masing-masing *cluster* tampak lebih merata. K-Medoids berhasil memisahkan *cluster* berdasarkan pola Masa Pajak secara lebih jelas, di mana *cluster* 1 memiliki nilai masa pajak paling tinggi dan stabil, *cluster* 0 berada pada rentang menengah, dan *cluster* 2 pada rentang rendah ke menengah. Tidak terdapat *cluster* dengan anggota sangat sedikit, menunjukkan bahwa K-Medoids lebih *robust* terhadap *outlier*. Hasil ini konsisten dengan nilai DBI yang menunjukkan bahwa K-Means memiliki kualitas *cluster* lebih baik karena menghasilkan nilai DBI yang lebih rendah (DBI = 0.380) dibanding K-Medoids (DBI = 1.141).

4.8. Perbandingan dan Analisis

Berdasarkan hasil pengelompokan dengan metode K-Means dan K-Medoids, dapat dilakukan beberapa perbandingan untuk melihat metode mana yang memberikan hasil lebih baik terhadap data faktur pajak *CORETAX*.

a. Aspek nilai (*Davies–Bouldin Index*) DBI, metode K-Means menunjukkan performa yang lebih baik karena menghasilkan nilai DBI sebesar 0.380, lebih rendah dibandingkan K-Medoids yang memperoleh nilai 1.141. Nilai DBI yang lebih rendah menandakan bahwa *cluster* yang terbentuk lebih kompak dan memiliki pemisahan yang lebih jelas antar *cluster*, sehingga kualitas pengelompokan K-Medoids lebih unggul.
b. Sensitivitas terhadap *outlier*, K-Means terlihat jauh lebih sensitif. Hal ini ditunjukkan dari munculnya satu *cluster* kecil yang hanya berisi dua data akibat adanya nilai ekstrim. Sebaliknya, K-Medoids lebih tahan terhadap *outlier* karena menggunakan objek aktual sebagai pusat *cluster*, sehingga hasil pengelompokan lebih stabil dan tidak mudah terpengaruh oleh data ekstrim.
c. Distribusi anggota *cluster*, hasil K-Means menunjukkan adanya ketidakseimbangan dengan terbentuknya *cluster* berukuran sangat kecil. Pada metode K-Medoids, distribusi data antar *cluster* jauh lebih merata sehingga lebih representatif dalam menggambarkan karakteristik keseluruhan dataset.
d. Stabilitas hasil, K-Medoids memberikan struktur *cluster* yang lebih konsisten, sementara K-Means menghasilkan pemisahan *cluster* yang cenderung berubah dan kurang stabil ketika terdapat variasi nilai yang jauh berbeda.

Secara keseluruhan, berdasarkan nilai DBI ketahanan terhadap *outlier*, pemerataan jumlah anggota *cluster*, dan tingkat stabilitas hasil, metode K-Medoids menunjukkan performa yang lebih baik dibandingkan K-Means dalam pengelompokan data faktur pajak *CORETAX* di PT Agung Servis Jaya Sakti

5. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis dan pembahasan penelitian, dapat disimpulkan bahwa penerapan metode *clustering* menggunakan algoritma K-Means dan K-Medoids pada data faktur pajak elektronik berbasis sistem *CORETAX* di PT Agung Servis Jaya Sakti berhasil mengelompokkan pola transaksi perpajakan pelanggan secara efektif berdasarkan variabel DPP, PPN, dan Masa Pajak yang telah melalui tahapan data *selection*, data *cleaning*, transformasi, dan normalisasi. Kedua metode menghasilkan tiga *cluster* utama yang merepresentasikan transaksi dengan nilai pajak rendah, sedang, dan tinggi. Namun hasil evaluasi menggunakan *Davies–Bouldin Index* (DBI) menunjukkan bahwa metode K-Means memiliki nilai DBI terendah pada jumlah *cluster* K = 3 sehingga dinilai lebih optimal dalam menghasilkan *cluster* yang kompak dan terpisah dengan baik dibandingkan K-Medoids. Meskipun demikian, metode K-Medoids tetap relevan digunakan karena lebih stabil terhadap variasi data dan *outlier*. Secara keseluruhan, hasil *clustering* yang diperoleh mampu memberikan gambaran segmentasi pelanggan dan pola transaksi perpajakan secara sistematis, sehingga dapat dimanfaatkan sebagai dasar pendukung pengambilan keputusan dalam pengelolaan administrasi perpajakan, pemantauan transaksi, serta penyusunan strategi bisnis yang lebih efektif dan berbasis data.

DAFTAR PUSTAKA

[1] B. Kristanto, A. Turmudi Zy, and M. Fatchan, "Analisis Penentuan Karyawan Tetap Dengan Algoritma K-Means Dan Davies Bouldin Index," *Bull. Inf. Technol.*, vol. 4, no. 1, pp. 112–120, 2023, doi: 10.47065/bit.v4i1.521.

[2] G. B. Kaligis and S. Yulianto, "ANALISA PERBANDINGAN ALGORITMA K-MEANS, K-MEDOIDS, DAN X-MEANS UNTUK PENGELOMPOKKAN KINERJA PEGAWAI (Studi Kasus: Sekretariat DPRD Provinsi Sulawesi Utara)," *IT-EXPLORE*, vol. 01, no. 3, pp. 179–193, 2022.

[3] E. Purwaningsih and E. Nurelasari, "Implementasi Metode K-Means Clustering Dengan Davies Bouldin Index Pada Analisis Faktor Penyebab Perceraian," *Inf. Manag. Educ. Prof.*, vol. 7, no. 2, pp. 134–143, 2023, [Online]. Available: https://databoks.katadata.co.id/

[4] M. Wahyudi, S. Solikhun, and L. Pujiastuti, "Komparasi K-Means Clustering dan K-Medoids Clustering dalam Mengelompokkan Produksi Susu Segar di Indonesia Berdasarkan Nilai DBI," *J. Bumigora Inf. Technol.*, vol. 4, no. 2, pp. 243–254, 2022, doi: 10.30812/bite.v4i2.2104.

[5] R. Gustrianda and D. I. Mulyana, "Penerapan Data Mining Dalam Pemilihan Produk Unggulan dengan Metode Algoritma K-Means Dan K-Medoids," *J. Media Inform. Budidarma*, vol. 6, no. 1, pp. 27–34, 2022, doi: 10.30865/mib.v6i1.3294.

[6] B. G. Sudarsono, M. I. Leo, A. Santoso, and F. Hendrawan4, "Analysis Data Mining Netflix Data Using The Rapid Miner Application," *J. Bus. Audit Inf. Syst.*, vol. 4, no. 1, pp. 13–21, 2021.

[7] Idham, H. Rosika, and Yuliadi, "IMPLEMENTASI RAPIDMINER UNTUK CLESTERING DATA PENJUALAN PAKAIAN MENGGUNAKAN METODE K-MEANS," *J. Educ. Technol.*, vol. 5, no. 1, pp. 1–11, 2024.

[8] E. Tasia and M. Afdal, "Perbandingan Algoritma K-Means Dan K-Medoids Untuk Clustering Daerah Rawan Banjir Di Kabupaten Rokan Hilir," *Indones. J. Inform. Res. Softw. Eng.*, vol. 3, no. 1, pp. 65–73, 2023, doi: 10.57152/ijirse.v3i1.523.

[9] A. Ikhram and Sani Mutia, "Penerapan Algoritma K-Medoids untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Kemiskinan: Akses Perumahan dan Kesejahteraan," *Emerg. Stat. Data Sci. J.*, vol. 3, no. 2, pp. 583–596, 2025, doi: 10.20885/esds.vol3.iss.2.art13.

[10] G. N. Wala and R. Tesalonika, "Transformasi Administrasi Perpajakan Melalui Coretax: Analisis Hukum dan Akuntansi," *J. Komun. dan Ilmu Sos.*, vol. 2, no. 4, pp. 149–158, 2024, doi: 10.38035/jkis.v2i4.1479.

[11] M. D. Chandra, E. Irawan, I. S. Saragih, A. P. Windarto, and D. Suhendro, "Penerapan Algoritma K-Means dalam Mengelompokkan Balita yang Mengalami Gizi Buruk Menurut Provinsi," *BIOS J. Teknol. Inf. dan Rekayasa Komput.*, vol. 2, no. 1, pp. 30–38, 2021, doi: 10.37148/bios.v2i1.19.

[12] F. Naufal, Y. H. Chrisnanto, and A. K. Ningsih, "Sistem Rekomendasi Penawaran Produk Pada Online Shop Menggunakan K-Means Clustering," *Informatics Digit. Expert*, vol. 4, no. 1, pp. 10–17, 2023, doi: 10.36423/index.v4i1.879.

[13] S. Bahri and D. M. Midyanti, "Penerapan Metode K-Medoids untuk Pengelompokan Mahasiswa Berpotensi Drop Out," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 1, pp. 165–172, 2023, doi: 10.25126/jtiik.2023106643.

[14] Y. Syahra, D. R. Habibie, M. Nasution, and H. N. Nasution, "Klasterisasi Data Penanganan Dan Pelayanan Kesehatan Masyarakat Menggunakan Algoritma K-Means," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 9, no. 5, pp. 1423–1433, 2022, doi: 10.53513/jursi.v1i3.5223.