

PENERAPAN ALGORITMA K-MEANS CLUSTERING DATA TRANSAKSI ALAT BERAT DI CV ABC

Fiolin Meldy Graciella, Dien Novita

Sistem Informasi, Universitas Multi Data Palembang

Jl. Rajawali No. 14, Palembang, Indonesia

fiolinmeldygraciella_2226240061@mhs.mdp.ac.id

ABSTRAK

CV ABC memiliki data transaksi penyewaan alat berat yang terus bertambah dalam beberapa periode. Namun, data tersebut belum dimanfaatkan secara optimal untuk mengidentifikasi transaksi seperti apa yang mempengaruhi pendapatan perusahaan. Penelitian ini bertujuan untuk mengelompokkan data transaksi dengan jumlah cluster yang optimal sehingga dapat memberikan pemahaman kepada perusahaan terhadap karakteristik transaksi. Metode penelitian yang dilakukan adalah identifikasi masalah, pengumpulan data, seleksi data, pembersihan data, transformasi data menggunakan tools RapidMiner, pemrosesan data dengan algoritma K-Means, evaluasi model, serta interpretasi hasil dengan bantuan dashboard. Penentuan jumlah cluster optimal dilakukan menggunakan Davies-Bouldin Index (DBI) dengan nilai DBI sebesar 0,771, sedangkan kualitas hasil pengelompokan dievaluasi menggunakan Silhouette Coefficient dan menghasilkan nilai 0,5596 yang menyatakan struktur cluster berada pada kategori menengah. Data yang digunakan berjumlah 521 transaksi penyewaan alat berat periode Desember 2021 hingga Desember 2025. Hasil penelitian menyatakan bahwa jumlah cluster optimal adalah tiga cluster, yaitu cluster 1 (transaksi rendah) terdiri dari 108 data, cluster 2 (transaksi sedang) terdiri dari 50 data, dan cluster 3 (transaksi tinggi) terdiri dari 363 data. Kesimpulannya, penerapan algoritma K-Means dalam pengelompokan data transaksi mampu membantu perusahaan dalam pengambilan keputusan berdasarkan data, khususnya dalam mengevaluasi penggunaan alat berat, penentuan strategi pengadaan, dan penyesuaian harga sewa alat berat.

Kata kunci : Alat berat, Data Mining, Clustering, K-Means.

1. PENDAHULUAN

Perkembangan teknologi informasi telah mendorong berbagai perusahaan untuk memanfaatkan data sebagai sumber informasi dalam mendukung pengambilan keputusan bisnis. Seiring meningkatnya aktivitas operasional perusahaan, jumlah data yang dihasilkan juga semakin besar dan kompleks. Data tersebut tidak hanya berfungsi sebagai arsip administrasi, tetapi juga dapat dimanfaatkan untuk menemukan informasi yang bernilai bagi perusahaan. Salah satu pendekatan yang dapat digunakan untuk menggali informasi dari data dalam jumlah besar adalah *data mining*. *Data mining* adalah teknik dalam mencari pola yang terselubung dari sekumpulan data yang besar untuk menghasilkan wawasan baru, berharga, dan bernilai [1].

CV ABC merupakan perusahaan yang bergerak di bidang penyewaan alat berat dan menyediakan berbagai jenis alat untuk mendukung kebutuhan proyek konstruksi maupun pekerjaan sipil. Perusahaan ini menyediakan berbagai jenis alat berat untuk beragam kebutuhan proyek, seperti *excavator*, *bulldozer*, dan *vibro* yang memiliki variasi tipe. Dalam kegiatan operasionalnya, perusahaan menghasilkan data transaksi penyewaan alat berat dalam jumlah yang cukup besar yang digunakan sebagai arsip administrasi dan pelaporan pendapatan perusahaan. Meskipun demikian, data transaksi tersebut belum dianalisis secara mendalam sehingga perusahaan belum memiliki informasi mengenai karakteristik transaksi penyewaan yang berkontribusi terhadap peningkatan

maupun penurunan pendapatan. Akibatnya, perusahaan mengalami kesulitan dalam melakukan evaluasi penggunaan alat berat, menentukan strategi pengadaan unit, serta menyusun kebijakan bisnis yang didasarkan pada informasi yang objektif.

Salah satu teknik dalam *data mining* yang dapat digunakan untuk mengidentifikasi pola pada data transaksi adalah *clustering*. *Clustering* merupakan teknik untuk mengenali pola kelompok dalam sebuah kumpulan data. Karakteristik data dalam satu kelompok cenderung memiliki kemiripan yang lebih tinggi dibandingkan dengan data yang berada di kelompok berbeda [2]. Dalam mengelompokkan transaksi penyewaan, digunakan algoritma yang bernama K-Means. Algoritma K-Means adalah algoritma *clustering* yang menggunakan parameter *k* sebagai jumlah *cluster* yang ingin dibentuk. Prosesnya mengelompokkan banyaknya data ke dalam *k cluster* dengan prinsip bahwa data dalam satu *cluster* memiliki karakteristik yang serupa, sementara data antar *cluster* lain memiliki tingkat kesamaan yang rendah [3]. Keunggulan dari penerapan algoritma K-Means adalah memiliki kemampuan untuk mengelompokkan data secara otomatis berdasarkan kesamaan pola transaksi sehingga sistem dapat membentuk karakteristik secara objektif menjadi tiga kelompok utama, yakni kategori rendah, sedang, dan tinggi [4]. Algoritma K-Means termasuk salah satu metode pengelompokan data yang sederhana serta banyak diterapkan dalam berbagai penelitian. Metode ini sering digunakan karena mampu menangani data dalam jumlah besar dan

melakukan proses pengelompokan dengan waktu komputasi yang cepat serta efisien [5]. Dalam memperoleh jumlah *cluster* yang optimal digunakan metode Davies-Bouldin Index (DBI), sedangkan kualitas *cluster* yang dihasilkan dievaluasi menggunakan Silhouette Coefficient. Kombinasi kedua metode evaluasi tersebut diharapkan mampu menghasilkan pengelompokan yang lebih representatif sesuai dengan karakteristik data transaksi penyewaan alat berat yang dimiliki perusahaan. Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengelompokkan data transaksi penyewaan alat berat pada CV ABC berdasarkan karakteristik transaksi yang dimiliki. Hasil pengelompokan diharapkan dapat memberikan informasi mengenai kelompok transaksi dengan tingkat aktivitas yang berbeda sehingga dapat membantu perusahaan dalam melakukan evaluasi penggunaan alat, menentukan strategi harga dan pengadaan alat, serta mendukung proses pengambilan keputusan bisnis yang lebih efektif dan berbasis data.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian terdahulu telah banyak menerapkan algoritma K-Means untuk pengelompokan data, seperti pada penelitian [4] dan [6]. Namun, penelitian tersebut menentukan jumlah *cluster* (k) di awal secara acak. Selain itu, penelitian tersebut juga berfokus pada *rental* mobil dan risiko Covid-19. Penelitian ini menerapkan pendekatan dengan menggunakan Davies-Bouldin Index (DBI) di awal tahapan, seperti pada penelitian [7] untuk menentukan jumlah *cluster* yang terbaik atau optimal. Setelah jumlah *cluster* terbaik diperoleh melalui pengujian DBI, kualitas pemisahan serta kekuatan struktur antar-*cluster* divalidasi menggunakan Silhouette Coefficient untuk menilai kualitas pengelompokan yang terbentuk, seperti pada pendekatan penelitian [8]. Hal ini diterapkan pada data transaksi penyewaan alat berat di CV ABC guna menghasilkan pengelompokan yang akurat dalam menganalisis karakteristik transaksi penyewaan alat berat yang memengaruhi peningkatan maupun penurunan pendapatan perusahaan.

2.2. Knowledge Discovery in Database (KDD)

Metode Knowledge Discovery in Database (KDD) terdiri atas lima tahapan utama [7]:

- a. *Data Selection*
Tahap awal adalah pemilihan data yang relevan terhadap tujuan penelitian.
- b. *Preprocessing*
Tahap ini adalah penghapusan data kosong (*null*), menghilangkan atribut yang tidak relevan, dan penanganan data yang mengandung kesalahan.
- c. *Transformation*
Berkfungsi untuk mengubah data mentah ke dalam bentuk yang sesuai dengan kebutuhan algoritma yang mencakup konversi data non-numerik menjadi numerik agar dapat diolah.

d. Data Mining

Tahap *data mining* merupakan inti dari keseluruhan proses untuk mencari pola atau hubungan tersembunyi dalam suatu data dengan algoritma tertentu yang memberikan informasi bagi pengambilan keputusan.

e. Evaluation

Tahapan untuk menilai pola yang telah ditemukan. Hasil evaluasi disajikan dalam bentuk visual, seperti tabel dan grafik agar lebih mudah dipahami.

2.3. Algoritma K-Means

Algoritma K-Means merupakan metode *clustering* yang digunakan untuk mengelompokkan data ke dalam beberapa *cluster* berdasarkan tingkat kemiripan karakteristik data. Prosesnya diawali dengan penentuan *centroid* awal, kemudian setiap data dikelompokkan ke *cluster* dengan jarak terdekat dan *centroid* diperbarui secara iteratif hingga mencapai kondisi *konvergen*. K-Means memiliki keunggulan dalam mengolah data berukuran besar dengan proses yang relatif cepat dan efisien. Namun, hasil pengelompokan dipengaruhi oleh pemilihan *centroid* awal dan jumlah *cluster* yang ditentukan sebelumnya. Oleh karena itu, penentuan jumlah *cluster* yang optimal menjadi faktor penting untuk menghasilkan kualitas *clustering* yang baik [9].

2.4. Davies-Bouldin Index (DBI)

DBI adalah sebuah metode yang digunakan untuk menilai seberapa optimal hasil pengelompokan dengan melihat nilai dari kohesi (kemiripan data di dalam satu *cluster*) dan separasi (jarak antar pusat *cluster*). Metode ini bertujuan untuk menghasilkan *cluster* yang memiliki tingkat kemiripan tinggi di dalam *cluster* dan perbedaan yang jelas antar *cluster* [10]. Apabila nilai Davies-Bouldin Index (DBI) semakin mendekati nol, maka hasil *cluster* yang terbentuk semakin optimal [11].

2.5. Silhouette Coefficient

Silhouette Coefficient merupakan metode yang digunakan untuk mengukur kualitas hasil pengelompokan data dengan mengidentifikasi tingkat kesesuaian setiap objek terhadap *cluster*. Nilai Silhouette Coefficient berada pada rentang -1 hingga 1.

Suatu hasil *clustering* dikatakan baik apabila menghasilkan nilai positif yang mendekati 1. Nilai yang mendekati 0 menunjukkan bahwa objek berada pada batas antar *cluster*, sedangkan nilai negatif menunjukkan bahwa objek kemungkinan ditempatkan pada *cluster* yang kurang tepat [12]. Tingkat kekuatan struktur *cluster* dievaluasi berdasarkan pengelompokan nilai koefisien *Silhouette*, yaitu [8]:

- a. Nilai $0,7 < s \leq 1$: Struktur kelompok kuat (*strong*).
- b. Nilai $0,5 < s \leq 0,7$: Struktur kelompok menengah (*medium*).

- c. Nilai $0,25 < s \leq 0,5$: Struktur kelompok lemah (*weak*).
- d. Nilai $s \leq 0,25$: Tidak membentuk struktur *cluster* yang signifikan (*no structure*).

3. METODE PENELITIAN

3.1. Identifikasi Masalah

Bertujuan untuk memahami kondisi dan permasalahan yang dihadapi. Berdasarkan hasil observasi dan wawancara, diketahui bahwa data transaksi penyewaan alat tercatat di Microsoft Excel sebagai pelaporan pendapatan. Namun, data transaksi belum dianalisis mendalam sehingga perusahaan belum mengetahui transaksi seperti apa yang membuat pendapatan mengalami peningkatan maupun penurunan.

3.2. Pengumpulan Data

Data dikumpulkan melalui tiga metode, yaitu observasi, wawancara, dan dokumentasi. Observasi dilakukan untuk memahami alur pencatatan transaksi penyewaan alat berat di perusahaan. Wawancara dilakukan dengan pihak perusahaan guna memperoleh informasi terkait proses bisnis dan permasalahan yang dihadapi. Dokumentasi dilakukan untuk mengumpulkan data historis transaksi penyewaan alat berat yang diinput di Microsoft Excel.

Data yang digunakan untuk penelitian, mencakup data transaksi penyewaan dari periode Desember 2021 hingga Desember 2025. Data transaksi penyewaan alat berat CV ABC memiliki 521 baris data yang terdiri dari delapan atribut, seperti tanggal, Surat Perintah Kerja (SPK), Pelanggan, Lokasi Pekerjaan, Jenis Alat, Jam Kerja, Harga Satuan, dan Total Harga.

3.3. Seleksi Data

Pada tahap seleksi data dilakukan pemilihan atribut yang relevan. Atribut yang digunakan pada penelitian ini terdiri dari Lokasi Pekerjaan, Jenis Alat, Jam Kerja, dan Total Harga. Pemilihan atribut tersebut didasarkan pada informasi utama yang terdapat dalam data transaksi penyewaan alat berat serta relevansi terhadap tujuan penelitian.

3.4. Pembersihan Data

Tahap pembersihan data dilakukan untuk memastikan kualitas *dataset* sebelum diproses. Proses ini meliputi pengecekan data kosong, data duplikat, dan inkonsistensi data. Selain itu dilakukan penyesuaian format data agar seluruh atribut memiliki struktur yang konsisten dan siap digunakan pada proses *clustering*.

3.5. Transformasi Data

Tahap transformasi dilakukan dengan mengubah atribut kategorikal menjadi numerik dengan bantuan *software* RapidMiner. RapidMiner merupakan perangkat lunak yang digunakan untuk melakukan pengolahan dan analisis data, khususnya dalam bidang *data mining* [7]. Selanjutnya dilakukan normalisasi

data agar setiap atribut memiliki rentang nilai yang seimbang sehingga proses perhitungan jarak pada algoritma K-Means dapat berjalan secara optimal.

3.6. Proses Data Mining

Data yang telah ditransformasi tersebut diproses menggunakan algoritma K-Means. Proses *clustering* dilakukan dengan menentukan jumlah *cluster* yang optimal dengan menggunakan metode Davies-Bouldin Index (DBI), menentukan *centroid* awal secara acak, menghitung jarak antar data menggunakan *Euclidean Distance*, melakukan pengelompokan data, dan memperbarui *centroid* secara iteratif hingga mencapai kondisi konvergen.

3.7. Evaluasi

Evaluasi model dilakukan menggunakan metode Silhouette Coefficient untuk menilai *cluster* yang telah terbentuk.

3.8. Interpretasi Hasil

Tahap interpretasi dilakukan dengan menganalisis karakteristik setiap *cluster* yang terbentuk sehingga dapat memberikan informasi mengenai kelompok transaksi penyewaan alat berat dengan kategori rendah, sedang, dan tinggi. Hasil *cluster* disajikan dalam bentuk *dashboard* yang mudah dipahami.

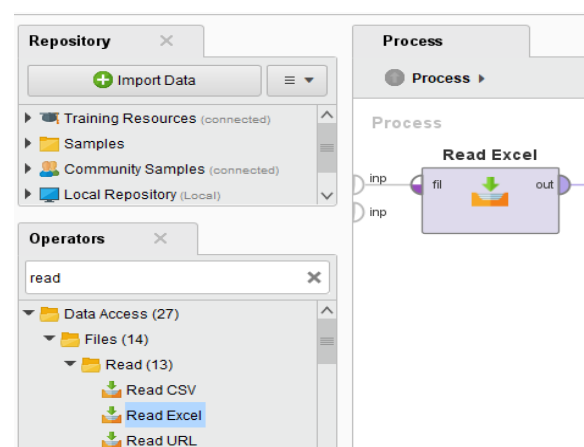
4. HASIL DAN PEMBAHASAN

4.1. Implementasi Clustering K-Means pada Transaksi Penyewaan Alat Berat

Berikut merupakan tahapan dalam pemilihan dan pemrosesan data menggunakan algoritma K-Means:

4.2. Seleksi Data

Proses pemilihan data dibantu dengan menggunakan *software* RapidMiner.



Gambar 1. Operator Read Excel

Pada Gambar 1, *dataset* transaksi dibaca menggunakan operator *Read Excel*, sedangkan untuk seleksi atribut, dengan cara *exclude column* terhadap atribut yang tidak digunakan untuk proses *clustering*, seperti Tanggal, SPK, Pelanggan, dan Harga Satuan.

Tabel 1. Hasil Seleksi Data

No	Lokasi Pekerjaan	Jenis Alat	Jam Kerja	Total Harga
1	STA Muaralawai	Vibrator Roller	200	Rp 46.000.000
2	STA Muaralawai	Excavator PC200	200	Rp 46.000.000
3	STA Muaralawai	Excavator PC200	200	Rp 48.000.000
...	...	...	...	...
470	Palembang	Vibrator Roller	1 Hari	Rp 3.500.000
471	Palembang	Excavator PC78	2 Hari	Rp 5.400.000
...	...	...	...	...
514	Merapi-Divre III	Excavator PC70	200	Rp 44.000.000
515	Tanjung Enim Baru	Excavator PC200	200	Rp 56.000.000
516	Tanjung Enim Baru	Excavator PC200	200	Rp 56.000.000
517	Kramasan Divre III	Excavator PC75	100	Rp 22.000.000
518	STA Merapi	Excavator PC200	200	Rp 50.000.000
519	STA Banjarsari-Merapi	Bulldozer D6	200	Rp 65.000.000
520	STA Banjarsari-Merapi	Excavator PC200	200	Rp 50.000.000
521	STA Banjarsari-Merapi	Excavator PC75	156	Rp 34.320.000

Tabel 1 merupakan tabel hasil seleksi data yang telah dilakukan pemilihan atribut relevan dan siap untuk dilakukan pembersihan data.

4.3. Pembersihan Data

Pada tahap pembersihan data, data transaksi penyewaan tidak terdapat data kosong (*null*) maupun data ganda (duplikat) sehingga *dataset* transaksi tetap berjumlah 521 baris.

Tabel 2. Hasil Pembersihan Data

No	Lokasi Pekerjaan	Jenis Alat	Jam Kerja	Total Harga
1	STA Muaralawai	Vibrator Roller	200	46000000
2	STA Muaralawai	Excavator PC200	200	46000000
3	STA Muaralawai	Excavator PC200	200	48000000
...	...	...	...	...
470	Palembang	Vibrator Roller	24	3500000
471	Palembang	Excavator PC78	48	5400000
...	...	...	...	...
514	Merapi-Divre III	Excavator PC70	200	44000000
515	Tanjung Enim Baru	Excavator PC200	200	56000000
516	Tanjung Enim Baru	Excavator PC200	200	56000000
517	Kramasan Divre III	Excavator PC75	100	22000000
518	STA Merapi	Excavator PC200	200	50000000
519	STA Banjarsari-Merapi	Bulldozer D6	200	65000000
520	STA Banjarsari-Merapi	Excavator PC200	200	50000000
521	STA Banjarsari-Merapi	Excavator PC75	156	34320000

Pada Tabel 2, pembersihan data dilakukan untuk menjaga konsistensi dengan mengonversi atribut Jam Kerja dalam bentuk hari, seperti '1 Hari' dan '2 Hari', menjadi satuan jam yaitu '24' dan '48' (*integer*).

Selain itu, dilakukan penghapusan simbol mata uang 'Rp' pada atribut Total Harga untuk menyeragamkan format menjadi data kuantitatif agar dapat diproses oleh algoritma K-Means.

Tabel 3. Hasil Transformasi dan Normalisasi Data

No	Lokasi Pekerjaan	Jenis Alat	Jam Kerja	Total Harga
1	-1,171	-0,868	0,446	0,163
2	-1,171	-0,489	0,446	0,163
3	-1,171	-0,489	0,446	0,313
...	...	...	...	...
470	0,454	1,404	-2,586	-1,591
471	-0,876	-0,489	0,446	0,462
...	...	...	...	...
514	1,192	6,706	0,446	0,014
515	1,931	-0,489	0,446	0,910
516	1,931	-0,489	0,446	0,910
517	2,152	0,268	-1,575	-1,629
518	1,488	-0,489	0,446	0,462
519	0,232	-0,110	0,446	1,582
520	0,232	-0,489	0,446	0,462
521	0,232	0,268	-0,443	-0,709

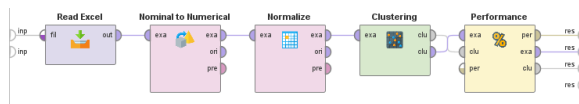
Tabel 3 merupakan *dataset* yang telah dilakukan transformasi dan normalisasi. Normalisasi dilakukan dengan menyesuaikan setiap nilai data terhadap rata-rata atributnya sehingga nilai akhir berada di sekitar angka nol. Nilai positif menunjukkan bahwa data berada di atas rata-rata, sedangkan nilai negatif menunjukkan bahwa data berada di bawah rata-rata. Sementara itu, nilai yang mendekati nol menunjukkan bahwa data tersebut berada di sekitar rata-rata.

4.4. Proses Data Mining

Langkah selanjutnya adalah melakukan proses *data mining*. Berikut adalah beberapa tahapan algoritma K-Means, antara lain [4]:

- Menentukan jumlah *cluster* (k) yang akan dibentuk.

Dalam menentukan jumlah *cluster* yang terbentuk secara optimal, dilakukan perhitungan menggunakan metode Davies-Bouldin Index (DBI) sesuai dengan struktur data transaksi penyewaan alat berat CV ABC. Sebelum itu, dilakukan rancangan model *clustering* untuk mendapatkan hasil perhitungan DBI.



Gambar 2. Rancangan Model *Clustering* K-Means

Gambar 2 adalah operator yang digunakan untuk proses k-means. Pengaturan *parameters* pada operator *Clustering* (K-Means) dilakukan dengan menetapkan $k = 2$ untuk melihat hasil DBI. Pada bagian *measure types*, dipilih opsi *Mixed Measures* karena *dataset* penelitian memiliki tipe data numerik dan nominal. Sedangkan pada jenis *mixed measure*, dipilih opsi *MixedEuclideanDistance*. Dalam mengukur validitas dari *cluster* yang terbentuk, digunakan operator *Performance* untuk melihat hasil perhitungan Davies-

Bouldin Index (DBI). Setelah menentukan jumlah *cluster* pada *parameters* di operator *Clustering* K-Means, hasil perhitungan Davies-Bouldin Index (DBI) dapat dilihat di tab *Results* (*PerformanceVector*).

Tabel 4. Tabel Perbandingan Nilai DBI

Jumlah Cluster (k)	Nilai DBI
2	0,88
3	0,771
4	0,773
5	0,83

Tabel 4 merupakan perbandingan nilai DBI untuk jumlah *cluster* $k = 2$ sampai $k = 5$. Perbandingan ini dilakukan untuk menemukan nilai DBI terkecil. Berdasarkan hasil perhitungan, nilai DBI terkecil adalah 0,771 sehingga jumlah *cluster* optimal untuk *clustering* transaksi penyewaan alat berat di CV ABC adalah $k = 3$.

Menetapkan titik pusat awal (*centroid*) secara acak sesuai dengan banyaknya *cluster*.

Tabel 5. Penentuan *Centroid* Acak

Baris Centroid	Lokasi Pekerjaan	Jenis Alat	Jam Kerja	Total Harga
C1	0,232	-0,489	0,446	0,462
C2	1,192	6,706	0,446	0,014
C3	1,931	-0,489	0,446	0,910

Tabel 5 adalah menentukan *centroid* acak berdasarkan banyaknya *cluster* yang ditentukan, sehingga:

C1 (0,232; -0,489; 0,446; 0,462)

C2 (1,192; 6,706; 0,446; 0,014)

C3 (1,931; -0,489; 0,446; 0,910)

- Menghitung jarak setiap data ke *centroid*. Perhitungan menggunakan rumus *Euclidean Distance* adalah sebagai berikut:

$$De = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (1)$$

Tabel 6. Hasil Perhitungan Jarak Setiap Data ke *Centroid*

No	Lokasi Pekerjaan	Jenis Alat	Jam Kerja	Total Harga	C1	C2	C3
514	1,192	6,706	0,446	0,014	7,27	0	7,29
515	1,931	-0,489	0,446	0,91	1,76	7,29	0
516	1,931	-0,489	0,446	0,91	1,76	7,29	0
517	2,152	0,268	-1,575	-1,629	3,57	7,01	3,34
518	1,488	-0,489	0,446	0,462	1,26	7,25	0,63
519	0,232	-0,11	0,446	1,582	1,18	7,06	1,87
520	0,232	-0,489	0,446	0,462	0	7,27	1,76
521	0,232	0,268	-0,443	-0,709	1,65	6,61	2,62

Tabel 6 merupakan hasil perhitungan yang dilakukan pada data ke-514 sampai data ke-521.

- Mengelompokkan setiap data pada *cluster* dengan jarak terdekat terhadap *centroid*.

Tabel 7. Pengelompokan Data Berdasarkan Hasil Perhitungan

No	Lokasi Pekerjaan	Jenis Alat	Jam Kerja	Total Harga	C1	C2	C3	Jarak Terdekat	Cluster
514	1,192	6,706	0,446	0,014	7,27	0	7,29	0	2
515	1,931	-0,489	0,446	0,91	1,76	7,29	0	0	3
516	1,931	-0,489	0,446	0,91	1,76	7,29	0	0	3
517	2,152	0,268	-1,575	-1,629	3,57	7,01	3,34	3,34	3
518	1,488	-0,489	0,446	0,462	1,26	7,25	0,63	0,63	3
519	0,232	-0,11	0,446	1,582	1,18	7,06	1,87	1,18	1
520	0,232	-0,489	0,446	0,462	0	7,27	1,76	0	1
521	0,232	0,268	-0,443	-0,709	1,65	6,61	2,62	1,65	1

Tabel 7 adalah mengelompokkan data dengan jarak terdekat terhadap *centroid* berdasarkan hasil perhitungan yang telah dilakukan.

- d. Menghitung ulang posisi *centroid* berdasarkan rata-rata data dalam setiap *cluster* dengan rumus:

$$c_k = \frac{1}{n_k} \sum d_i \quad (2)$$

- e. Mengulangi proses hingga posisi *centroid* tidak berubah lagi (telah *konvergen*).

Setelah mendapatkan nilai *centroid* baru, dilakukan pengulangan proses dari tahap ke-3 untuk perhitungan jarak dan pembentukan *cluster* menggunakan *centroid* baru. Proses ini terus dilakukan hingga tidak terjadi perubahan posisi *centroid* atau perpindahan data antar *cluster* (*konvergen*).

4.5. Evaluasi

Dalam menentukan jumlah *cluster* optimal, digunakan metode Davies-Bouldin Index (DBI) dan diperoleh nilai DBI terkecil pada $k = 3$ sehingga jumlah *cluster* yang digunakan pada proses *clustering* adalah tiga *cluster*. Selanjutnya, untuk memastikan kualitas model *clustering*, dilakukan evaluasi menggunakan Silhouette Coefficient.

Tabel 8. Hasil Perhitungan Silhouette Coefficient

Cluster	Nilai Silhouette
2	0,485
3	0,559
4	0,569
5	0,58

Tabel 8 adalah hasil perhitungan Silhouette Coefficient menggunakan Google Colab. Hasil evaluasi menunjukkan bahwa $k = 3$ memperoleh nilai silhouette sebesar 0,5596. Nilai tersebut berada pada rentang 0,5 – 0,7, yang menyatakan bahwa *cluster* termasuk struktur kelompok menengah (*medium*) sehingga model *clustering* dengan tiga *cluster* dinilai memiliki kualitas pengelompokan yang baik. Meskipun pada pengujian beberapa skenario *cluster* nilai silhouette tertinggi diperoleh pada $k = 5$, penelitian ini tetap menggunakan $k = 3$ mengacu pada hasil optimal Davies-Bouldin Index sebagai kriteria utama, sedangkan nilai silhouette berfungsi mengevaluasi kualitas model.

4.6. Interpretasi Hasil

Proses *clustering* k-means dianalisis menggunakan *software* RapidMiner dan berikut ini merupakan hasil dari *cluster* yang telah terbentuk:

Tabel 9. Hasil Data pada Cluster 1

No	Lokasi Pekerjaan	Jenis Alat	Jam Kerja	Total Harga	Cluster
4	STA Muaralawai	Bulldozer D6	100	Rp 31.000.000	Cluster 1
47	STA Muara Enim	Vibrator Roller	81	Rp 18.630.000	Cluster 1
49	Depo Simpang	Excavator Pc200	200	Rp 32.640.000	Cluster 1
...	...	...	...	...	...
502	Palembang	Excavator Pc78	96	Rp 12.000.000	Cluster 1
517	Kramasan Divre III	Excavator Pc75	100	Rp 22.000.000	Cluster 1
521	STA Banjarsari-Merapi	Excavator Pc75	156	Rp 34.320.000	Cluster 1

Tabel 9 merupakan data yang termasuk pada Cluster 1 (transaksi rendah) yang terdiri dari 108 data transaksi dengan rata-rata jam kerja terendah yaitu 85,44 jam dan rata-rata pendapatan sebesar Rp 18.067.870. Cluster ini mencerminkan penyewaan alat berat dengan durasi (jam kerja) relatif singkat dengan nilai harga sewa yang rendah. Berdasarkan hasil tersebut, perusahaan dapat melakukan evaluasi

terhadap kebijakan minimum penyewaan, seperti menetapkan batas minimal jam sewa, menerapkan deposit biaya mobilisasi alat, serta mempertimbangkan kenaikan tarif sewa harian untuk menutupi biaya operasional pada lokasi yang jauh dengan jam kerja yang singkat agar transaksi dengan durasi penyewaan yang singkat tetap memberikan kontribusi keuntungan bagi perusahaan.

Tabel 10. Hasil Data pada Cluster 2

No	Lokasi Pekerjaan	Jenis Alat	Jam Kerja	Total Harga	Cluster
14	STA Muaralawai	Bulldozer D6	200	Rp 62.000.000	Cluster 2
23	STA Muaralawai	Bulldozer D6	200	Rp 65.000.000	Cluster 2
56	Depo Simpang	Bulldozer D6	200	Rp 65.000.000	Cluster 2
...	...	...	...	...	...
515	Tanjung Enim Baru	Excavator Pc200	200	Rp 56.000.000	Cluster 2
516	Tanjung Enim Baru	Excavator Pc200	200	Rp 56.000.000	Cluster 2
519	STA Banjarsari-Merapi	Bulldozer D6	200	Rp 65.000.000	Cluster 2

Tabel 10 merupakan data yang termasuk pada Cluster 2 (transaksi sedang) yang terdiri dari 50 data transaksi. Cluster ini memiliki rata-rata jam kerja sebesar 193,56 jam dengan rata-rata pendapatan Rp 58.740.000. Kelompok ini memiliki nilai harga sewa per transaksi yang tinggi namun dengan frekuensi jumlah transaksi yang lebih sedikit dibandingkan cluster lainnya. Hal ini menunjukkan bahwa cluster ini

memiliki potensi keuntungan yang cukup besar pada setiap transaksi yang dilakukan, khususnya pada alat berat Bulldozer D6. Berdasarkan hasil tersebut, perusahaan dapat memfokuskan strategi bisnis pada jaminan kesiapan unit (*ready stock*), pemeliharaan kualitas unit, dan kenaikan harga sewa alat berat Bulldozer D6 untuk meningkatkan pendapatan perusahaan.

Tabel 11. Hasil Data pada Cluster 3

No	Lokasi Pekerjaan	Jenis Alat	Jam Kerja	Total Harga	Cluster
1	STA Muaralawai	Vibrator Roller	200	Rp 46.000.000	Cluster 3
2	STA Muaralawai	Excavator Pc200	200	Rp 46.000.000	Cluster 3
3	STA Muaralawai	Excavator Pc200	200	Rp 48.000.000	Cluster 3
...	...	...	...	...	...
514	Merapi-Divre III	Excavator Pc70	200	Rp 44.000.000	Cluster 3
518	STA Merapi	Excavator Pc200	200	Rp 50.000.000	Cluster 3
520	STA Banjarsari-Merapi	Excavator Pc200	200	Rp 50.000.000	Cluster 3

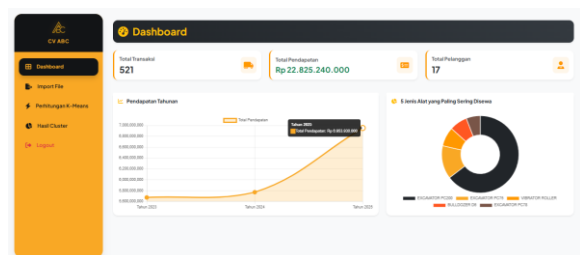
Tabel 11 merupakan data yang termasuk pada Cluster 3 (transaksi tinggi) yang terdiri dari 363 data transaksi. Cluster ini memiliki rata-rata jam kerja tertinggi yaitu 197,35 jam dan rata-rata pendapatan sebesar Rp 46.108.815. Cluster ini dikategorikan sebagai transaksi tinggi karena dominasinya dalam memberikan kontribusi pendapatan yang konsisten dan durasi pemakaian alat (jam kerja) yang maksimal bagi perusahaan, seperti sewa alat Excavator PC200. Hal tersebut menyatakan bahwa cluster ini merupakan kelompok pelanggan utama yang memberikan kontribusi pendapatan secara berkelanjutan bagi perusahaan. Berdasarkan hasil tersebut, perusahaan perlu menjaga kualitas layanan pada cluster ini melalui pemeliharaan unit secara berkala, memastikan ketersediaan alat berat, serta memberikan potongan harga sewa alat berat Excavator PC200 kepada pelanggan agar hubungan kerja sama jangka panjang dapat terus dipertahankan.

Pada tahap ini, hasil pengelompokan divisualisasikan dengan bantuan *dashboard* untuk mempermudah perusahaan dalam memahami kelompok transaksi. Adapun tampilan antarmuka dan fitur-fitur yang ada pada *dashboard* tersebut adalah sebagai berikut:

4.7. Menu Dashboard

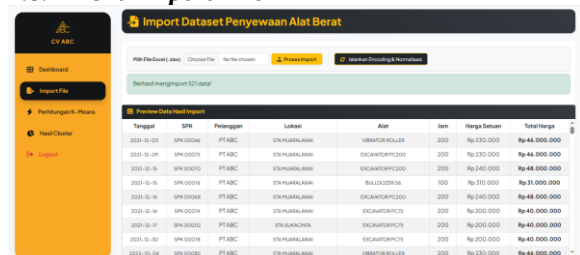
Gambar 3 merupakan tampilan menu *dashboard* yang menyajikan ringkasan informasi berupa total transaksi, total pendapatan, serta jumlah pelanggan. Selain itu, *dashboard* juga menampilkan visualisasi data dalam bentuk grafik yang menunjukkan

pendapatan tahunan dan jenis alat berat yang paling sering disewakan.



Gambar 3. Menu Dashboard

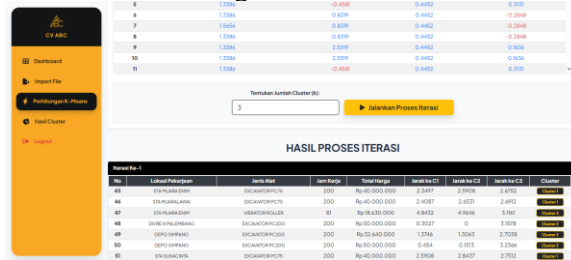
4.8. Menu Import File



Gambar 4. Menu Import File

Gambar 4 adalah tampilan menu *import file* agar pengguna dapat melakukan *import* data dalam format Excel yang kemudian akan ditampilkan sebagai *dataset* transaksi. Pengguna juga bisa menjalankan *encoding (nominal to numerical)* dan normalisasi dari data yang telah di *import*, lalu pengguna akan diarahkan ke menu perhitungan k-means.

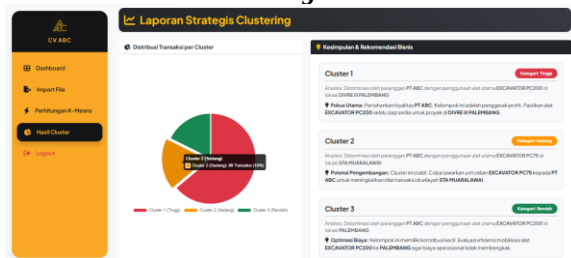
4.9. Menu Perhitungan K-Means



Gambar 5. Menu Perhitungan K-Means

Pada Gambar 5, data yang telah di *import* sebelumnya akan melalui proses *encoding* serta normalisasi pada atribut Lokasi Pekerjaan dan Jenis Alat. Setelah itu, dilakukan proses normalisasi pada atribut lokasi pekerjaan, jenis alat, jam kerja, dan total harga. Kemudian, pengguna menentukan jumlah *cluster* yang diinginkan. Maka, *dashboard* akan menampilkan hasil proses iterasi seperti pada Gambar 14. Pada Gambar 15, *dashboard* melakukan proses iterasi hingga mencapai kondisi *konvergen*. Lalu, pengguna dapat melihat hasil *clustering* dan visualisasi data.

4.10. Menu Hasil Clustering dan Visualisasi Data



Gambar 6. Menu Hasil Clustering dan Visualisasi Data

Data yang telah *konvergen* akan divisualisasikan dalam bentuk *pie chart*. Selain itu, *dashboard* juga menyajikan kesimpulan serta rekomendasi bisnis berdasarkan hasil pengelompokan transaksi yang telah terbentuk. Gambar 6 menampilkan *dataset* hasil *cluster* yang terbentuk berdasarkan *cluster* yang sudah ditentukan, yaitu $k = 3$.

5. KESIMPULAN DAN SARAN

Penerapan algoritma K-Means terhadap transaksi penyewaan alat berat terdapat sebanyak 521 baris data dan menggunakan empat atribut, yaitu Lokasi Pekerjaan, Jenis Alat, Jam Kerja, dan Total Harga. Dalam menentukan jumlah *cluster*, diperoleh jumlah *cluster* yang optimal menggunakan metode Davies-Bouldin Index (DBI), yaitu tiga *cluster*. *Cluster* 1 (transaksi rendah) sebanyak 108 data, *cluster* 2 (transaksi sedang) sebanyak 50 data, dan *cluster* 3 (transaksi tinggi) sebanyak 363 data. Selanjutnya, hasil *clustering* dievaluasi menggunakan metode Silhouette Coefficient dan diperoleh nilai sebesar 0,5596, yang menunjukkan bahwa *cluster* yang terbentuk berada pada struktur kelompok menengah

(medium). Dengan adanya hasil pengelompokan tersebut, perusahaan dapat lebih mudah mengidentifikasi kelompok transaksi sehingga dapat digunakan sebagai bahan evaluasi penggunaan alat berat untuk pengambilan keputusan bisnis berbasis data, seperti pengadaan alat berat, penentuan strategi harga sewa alat berat dan peningkatan kualitas layanan untuk pelanggan loyal. Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat diberikan untuk pengembangan penelitian selanjutnya. Penelitian selanjutnya diharapkan dapat melakukan perbandingan dengan algoritma lain untuk memperoleh hasil yang lebih baik. Selain itu, penambahan atribut data yang lebih lengkap dan jumlah data yang lebih besar dapat dilakukan agar hasil pengelompokan menjadi lebih akurat.

DAFTAR PUSTAKA

- [1] P. M. S. Tarigan, J. T. Hardinata, H. Qurniawan, M. Safii, and R. Winanjaya, "Implementasi data mining menggunakan algoritma apriori dalam menentukan persediaan barang (studi kasus: toko sinar harapan)," *J. Janitra Inform. dan Sist. Inf.*, vol. 2, no. 1, pp. 9–19, Apr. 2022, doi: 10.25008/janitra.v2i1.142.
- [2] B. Y. Putra, F. Y. Azzahra, and I. A. Erlanda, "Klasterisasi pengunjung mall menggunakan algoritma k-means berdasarkan pendapatan dan pengeluaran," *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 3, 2023.
- [3] M. Ikbali and E. Soemantri, "Implementasi data mining untuk menentukan strategi penjualan dengan metode k-means di cv.sentosa jaya," *J. Inform. dan Sist. Inf.*, vol. 15, no. 1, pp. 32–45, 2023.
- [4] S. W. Maesaroh, T. . Diansyah, R. Liza, and Y. F. A. Lubis, "Pemanfaatan algoritma k-means clustering pada sistem rental mobil," *J. Informatics Manag. Inf. Technol.*, vol. 5, no. 3, pp. 406–413, 2025, doi: 10.47065/jimat.v5i3.391.
- [5] Qomariyah and M. U. Siregar, "Comparative study of k-means clustering algorithm and k-medoids clustering in student data clustering," *J. Inform. Sunan Kalijaga*, vol. 7, no. 2, pp. 91–99, 2022.
- [6] Silviana, R. Kurniawan, A. Nazir, E. Budianita, F. Syafria, and S. K. Gusti, "Pengklasteran risiko covid-19 di riau menggunakan teknik one hot encoding dan algoritma k-means clustering," *J. Inf. dan Komput.*, vol. 10, no. 1, pp. 154–163, 2022, doi: 10.35959/jik.v10i1.291.
- [7] M. A. Aldi and Z. Fatah, "Implementasi k-means clustering dalam pengelompokan data kunjungan wisatawan asing di Indonesia," *J. Ilm. Multidisiplin Ilmu*, vol. 2, no. 1, pp. 13–19, 2025.
- [8] A. Atira and B. N. Sari, "Penerapan silhouette coefficient, elbow method dan gap statistics untuk penentuan cluster optimum dalam pengelompokan provinsi di indonesia

- berdasarkan indeks kebahagiaan,” *J. Ilm. Wahana Pendidik.*, vol. 9, no. 17, pp. 76–86, 2023.
- [9] E. Dwiguna and A. Bahtiar, “Penerapan data mining untuk menentukan penerima bantuan blt menggunakan metode clustering k-means pada desa pamulihan,” *J. Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1382–1388, 2024.
- [10] Irwan, W. Sanusi, A. S. Anwar, and A. Rahman, “The implementation of spatial model with k-means clustering method to cluster flood affected areas in bone regency,” *J. Soc. Sci. Humanit.*, vol. 3, no. 2, pp. 186–195, 2023.
- [11] N. A. Maori, “Metode elbow dalam optimasi jumlah cluster pada k-means clustering,” *J. SIMETRIS*, vol. 14, no. 2, pp. 277–287, 2023.
- [12] F. Dhewayani, D. Amelia, D. Nuralifah, B. N. Sari, and M. Jajuli, “Implementasi k-means clustering untuk pengelompokan daerah rawan bencana kebakaran menggunakan model crisp-dm,” *J. Teknol. dan Inf.*, vol. 12, no. 1, 2022, doi: 10.34010/jati.v12i1.