

# Implementasi Algoritma K-Means untuk Clustering Nominal Pembayaran pada Odoo Versi 11 di PT.XYZ

<sup>1)</sup> **Audrey Alicia Laseduw**

Program Studi Sistem Informasi, Universitas Pembangunan Jaya, Indonesia

E-Mail: [audrey.alicialaseduw@student.upj.ac.id](mailto:audrey.alicialaseduw@student.upj.ac.id)

<sup>2)</sup> **Cahyono Budy Santoso**

Program Studi Sistem Informasi, Universitas Pembangunan Jaya, Indonesia

E-Mail: [cahyono.budy@upj.ac.id](mailto:cahyono.budy@upj.ac.id)

## ABSTRACT

*Payment data in the Odoo version 11 system at PT. XYZ has not been optimally utilized to support strategic business decisions. This study implements the K-Means clustering algorithm to group customer payment amounts and identify transaction behavior patterns. The optimal number of clusters was determined using the Elbow Method and Silhouette Score, while visualization through scatter plot and block plot was used to interpret the clustering results. The analysis produced three main clusters representing customers with high, medium, and low payment amounts. These segmentation results enable PT. XYZ to better understand customer payment behavior, optimize marketing strategies, enhance service quality, and improve cash flow management. Overall, this research demonstrates the effectiveness of the K-Means algorithm in processing payment data within the Odoo ERP system and highlights its potential to support more accurate and data-driven business decision-making.*

**Keyword : ERP, Odoo, Data Mining, K-Means, Clustering**

## PENDAHULUAN

Perkembangan teknologi informasi yang pesat telah mendorong perusahaan untuk beradaptasi dalam mengelola data bisnis yang banyak. Salah satu sistem yang banyak digunakan untuk mendukung efisiensi operasional perusahaan adalah Enterprise Resource Planning (ERP), di mana seluruh proses bisnis terintegrasi dalam satu platform[1]. Odoo merupakan alternatif sistem ERP yang banyak diimplementasikan oleh perusahaan karena fleksibilitas dan kemampuannya dalam menyesuaikan kebutuhan dan menunjang efisiensi dan efektifitas pihak perusahaan juga pelanggan[2]. Namun, pengelolaan data transaksi yang besar di dalam Odoo sering kali belum dimanfaatkan secara optimal untuk analisis terkait nominal pembayaran pelanggan pada PT.XYZ.

PT. XYZ sebagai perusahaan yang telah mengimplementasikan Odoo versi 11 menghadapi permasalahan dalam menganalisis data nominal pembayaran pelanggan. Selama ini, data tersebut hanya disimpan sebagai arsip transaksi tanpa dilakukan analisis lebih lanjut untuk menemukan pola pembayaran pelanggan. Hal ini menyebabkan perusahaan kesulitan dalam menentukan strategi pemasaran atau kebijakan penagihan yang tepat. Oleh karena itu, perusahaan memiliki permintaan terkait suatu pendekatan analisis data untuk mengelompokkan pelanggan berdasarkan nominal pembayaran mereka.

Salah satu metode yang dapat digunakan untuk mengelompokkan data adalah data

mining, yaitu proses menggali informasi tersembunyi dari kumpulan data besar untuk mendukung pengambilan keputusan[3]. Dalam data mining, terdapat berbagai pendekatan algoritma yang digunakan untuk menemukan pola, salah satunya adalah algoritma unsupervised learning. Pendekatan unsupervised learning digunakan ketika data yang dianalisis tidak memiliki label atau kategori sebelumnya, sehingga sistem secara otomatis mencari struktur dan hubungan antar data[4]. Salah satu algoritma unsupervised learning yang populer adalah K-Means clustering, yang digunakan dalam penelitian kali ini bertujuan untuk mengklustering data berdasarkan tingkat kemiripan nilai antar objek.

Implementasi data mining menggunakan Algoritma K-Means bekerja dengan menganalisis dan menemukan pola sehingga menghasilkan klaster baru[5]. Metode ini banyak digunakan karena kesederhanaan, efisiensi komputasi, dan hasil pengelompokannya yang mudah dipahami. Dalam penelitian terhadap data yang ada di PT. XYZ, penerapan algoritma K-Means dapat membantu perusahaan dalam mengelompokkan pelanggan berdasarkan nominal pembayaran mereka, seperti pelanggan dengan nominal tinggi, sedang, dan rendah. Dengan adanya pengelompokan ini, perusahaan dapat memperoleh gambaran yang lebih jelas mengenai pola pembayaran pelanggan sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan strategis.

Penelitian ini bertujuan mengimplementasikan algoritma K-Means untuk clustering nominal pembayaran data yang telah diambil dari sistem Odoo versi 11 di PT. XYZ. Diharapkan hasil dari penelitian ialah dapat membantu perusahaan dalam pengambilan keputusan berbasis data, mengoptimalkan pengelolaan arus kas, menentukan prioritas layanan, serta merancang kebijakan penagihan yang lebih efektif dalam mengatur arus kas perusahaan, dan memberikan gambaran pola pembayaran pelanggan.

**METODE**

1. Metode Penelitian

Penelitian ini menggunakan metode kuantitatif yang berfokus pada pengolahan data numerik untuk menghasilkan pengelompokan pelanggan berdasarkan nominal pembayaran. Metode kuantitatif dipilih karena mampu memberikan hasil yang objektif dan terukur melalui analisis berbasis angka.

2. Sumber dan Jenis Data

a. Sumber Data

Data yang digunakan dalam penelitian ini merupakan data primer yang diperoleh secara langsung dari sistem informasi perusahaan, yaitu Odoo versi 11. Data primer dipilih karena memberikan informasi yang akurat dan sesuai dengan kondisi aktual perusahaan, sehingga

hasil analisis dapat mencerminkan pola pembayaran pelanggan secara nyata. Data yang diambil berupa nominal pembayaran pelanggan dari modul Accounting, sehingga sesuai untuk diolah menggunakan algoritma K-Means yang membutuhkan input berupa angka untuk proses pengelompokan. Pemilihan data primer ini juga bertujuan untuk memastikan bahwa hasil penelitian dapat digunakan sebagai dasar pengambilan keputusan strategis oleh perusahaan.

b. Jenis Data

Data yang digunakan pada penelitian ini berisi data pembayaran pelanggan dari rentang waktu 2020 hingga 2025 di PT.XYZ yang diekspor dari modul Accounting dengan format Microsoft Excel (xls).

**Gambar 1.** Data Pembayaran Pelanggan [sumber : PT.XYZ]

Dari data pembayaran pelanggan tersebut memiliki 10 atribut yang detail strukturnya dapat dilihat pada tabel dibawah ini.

**Tabel 1.** Deskripsi Data [sumber : Laseduw & Santoso, 2025]

No	Nama Atribut	Deskripsi	Tipe Data
1	Posted Date	Tanggal pembayaran diinput ke dalam sistem	Object
2	Name	Nomor Bank Masuk (BK) / Kas Masuk (KM)	Object
3	Payment Journal	Pembayaran masuk tertuju (rekening perusahaan)	Object
4	Payment Method Type	Metode Pembayaran	Object
5	Customer	Nama Customer	Object
6	Payment Amount	Nominal yang dibayarkan	Float64
7	Nomor Bukti Potong	Nomor / kode untuk mencatat bukti potong pajak	Object
8	Tanggal Bukti Potong	Tanggal diterbitkannya bukti potong pajak	Object
9	Status	Proses Approval oleh bagian keuangan	Object
10	Company	Pembayaran masuk ke perusahaan	Object

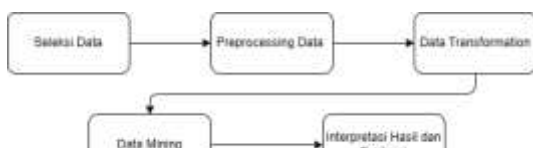
3. Teknik Pengolahan Data

Proses pengolahan data dalam penelitian ini dilakukan dengan pendekatan Knowledge Discovery in Database (KDD). Tahapan sistematis ini untuk menemukan pengetahuan yang berguna dari kumpulan data yang besar melalui serangkaian langkah analisis dan transformasi data[6].

Tahapan KDD pada penelitian ini meliputi seleksi data, pembersihan data, transformasi data, proses data mining, dan interpretasi hasil, seperti dijelaskan berikut ini:

a) Seleksi data

Pada tahap ini dilakukan pemilihan data yang relevan dengan tujuan penelitian. Data yang digunakan berasal dari sistem Odoo versi 11 yang diimplementasikan di PT. XYZ, khususnya data pembayaran pelanggan. Dari keseluruhan dataset yang tersedia, diambil atribut yang relevan secara langsung terhadap analisis pola pembayaran. Jumlah



**Gambar 2.** Teknik Pengolahan Data [sumber : Laseduw & Santoso, 2025]

data yang digunakan dalam penelitian ini adalah sekitar 10.429 transaksi pembayaran.

#### b) Preprocessing Data

Tahap preprocessing dilakukan untuk memastikan bahwa data yang digunakan dalam penelitian memiliki kualitas yang baik dan siap diolah menggunakan algoritma K-Means. Proses ini meliputi pembersihan data (*data cleaning*), penghapusan duplikasi, penanganan nilai hilang, serta transformasi/normalisasi untuk menyetarakan skala variabel numerik sebelum pemodelan[7]. Pada penelitian ini, proses data preprocessing dilakukan terhadap dataset pembayaran pelanggan yang diperoleh dari sistem Odoo versi 11 di PT. XYZ.

#### c) Data Transformation

Pada penelitian ini, atribut yang digunakan untuk analisis adalah Payment Amount, yaitu nominal pembayaran dari masing-masing pelanggan. Nilai nominal pembayaran memiliki rentang yang sangat bervariasi, jika tidak dilakukan transformasi, maka nilai dengan skala besar akan mendominasi hasil perhitungan jarak atau *euclidean distance* dan menyebabkan pembentukan cluster menjadi tidak seimbang. Maka, dilakukan proses normalisasi data menggunakan metode Min-Max Scaling[8].

#### d) Data Mining

Algoritma yang digunakan kali ini adalah K-Means Clustering, yang termasuk dalam kategori unsupervised learning[9]. Algoritma ini digunakan untuk mengelompokkan pelanggan berdasarkan kesamaan nominal pembayaran sehingga terbentuk beberapa kelompok atau kluster dengan karakteristik yang berbeda. Dilakukan metode Elbow atau *Elbow Method* dan *Silhouette Score* untuk mencari jumlah cluster yang optimal. Metode Elbow digunakan untuk mengidentifikasi titik optimal berdasarkan nilai inertia, sedangkan Silhouette Score digunakan untuk mengukur tingkat keseragaman antar data dalam cluster[10].

Setelah jumlah cluster ditentukan, algoritma K-Means dijalankan untuk melakukan proses klustering berdasarkan atribut Payment Amount yang telah dinormalisasi. Hasil dari tahap data mining ini berupa

pembentukan beberapa cluster pelanggan, yaitu kelompok dengan nominal pembayaran tinggi, sedang, dan rendah. Pola ini menjadi dasar dalam melakukan interpretasi lebih lanjut terhadap perilaku pembayaran pelanggan

#### e) Interpretasi Hasil dan Evaluasi

Tahap interpretasi hasil merupakan bagian akhir dari proses Knowledge Discovery in Database (KDD), yang bertujuan untuk memahami dan memberikan makna terhadap hasil clustering yang diperoleh dari algoritma K-Means. Setelah proses clustering dilakukan, terbentuk tiga kelompok utama pelanggan berdasarkan nominal pembayaran, yaitu cluster pelanggan dengan nominal pembayaran tinggi, sedang, dan rendah. Hasil pengelompokan ini divisualisasikan menggunakan scatter plot dan box plot, yang memudahkan dalam mengamati sebaran data setiap kluster dan mengamati sebaran data setiap kluster.

## HASIL DAN PEMBAHASAN

### 1. Seleksi Data

Tahap seleksi data dilakukan untuk memilih atribut atau variabel yang relevan dengan tujuan penelitian, yaitu mengelompokkan pelanggan berdasarkan nominal pembayaran yang dilakukan. Data yang digunakan berasal dari sistem Odoo versi 11 PT. XYZ yang berisi sepuluh atribut utama, yaitu: Posted Date, Name, Payment Journal, Payment Method Type, Customer, Payment Amount, Nomor Bukti Potong, Tanggal Bukti Potong, Status, dan Company.

Hal pertama yang dilakukan sebelum melakukan seleksi data ialah load dataset :

```
df = pd.read_excel("account.payment (2).xls")
print("Preview data:")
display(df.head())
```

Posted Date	Name	Payment Journal	Payment Method Type	Customer	Payment Amount	Nomor Bukti Potong	Tanggal Bukti Potong	Status	Company
2025-12-01 10:00:00	PT. XYZ	123456789	Bank Transfer	Customer A	1000000	BP001	2025-12-01	Setor	PT. XYZ
2025-12-02 11:00:00	PT. XYZ	987654321	Bank Transfer	Customer B	500000	BP002	2025-12-02	Setor	PT. XYZ
2025-12-03 09:00:00	PT. XYZ	234567890	Bank Transfer	Customer C	200000	BP003	2025-12-03	Setor	PT. XYZ
2025-12-04 12:00:00	PT. XYZ	345678901	Bank Transfer	Customer D	1500000	BP004	2025-12-04	Setor	PT. XYZ
2025-12-05 14:00:00	PT. XYZ	456789012	Bank Transfer	Customer E	750000	BP005	2025-12-05	Setor	PT. XYZ

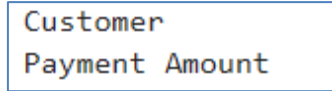
**Gambar 3.** Preview Data

[sumber : Laseduw & Santoso, 2025]

Setelah load data set berhasil dilakukan maka dilakukan proses seleksi atribut yang akan dilakukan sebelum implementasi algoritma K-

means, atribut yang dipilih ialah Customer & Payment Amount. Berikut perintah yang dilakukan untuk menjalankan seleksi data :

```
df = df[['Customer', 'Payment Amount']].copy()
```



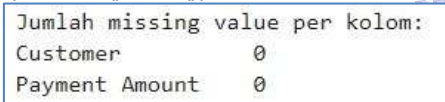
**Gambar 4.** Seleksi Data

[sumber : Laseduw & Santoso, 2025]

2. Preprocessing Data

Pada tahap ini proses data cleansing merupakan tahap penting agar kualitas analisis baik dan konsisten serta jauh dari kesalahan yang dapat mempengaruhi ke hasil klustering. Pengecekan *missing value* atau nilai kosong pada data menjadi hal pertama dalam data cleansing ini, perintah yang dijalankan untuk mengecek missing value adalah :

```
df = df[['Customer', 'Payment Amount']].copy()
print("Jumlah missing value per kolom:")
print(df.isnull().sum())
```



**Gambar 5.** Mengecek Missing Value Data  
 [sumber : Laseduw & Santoso, 2025]

Dilanjutkan dengan proses cleansing atau pembersihan data yang dimana diiringi dengan proses standarisasi format data pada atribut Payment Amount menjadi format numerik atau tipe data float64 agar dapat diproses benar oleh algoritma, berikut ditampilkan perintahnya :

```
df = df.drop_duplicates()
df = df.dropna(subset=['Customer', 'Payment Amount'])
df['Payment Amount'] = pd.to_numeric(df['Payment Amount'], errors='coerce')
df = df.dropna(subset=['Payment Amount'])
```



**Gambar 6.** Cleansing Data

[sumber : Laseduw & Santoso, 2025]

3. Data Transformasi

Proses normalisasi data dilakukan untuk menyetarakan skala nilai pada atribut Payment Amount agar berada dalam rentang 0 hingga 1, sehingga setiap data memiliki bobot yang seimbang dan tidak ada nilai yang mendominasi saat dilakukan proses clustering menggunakan algoritma K-Means. Kolom baru bernama "Payment\_Scaled" menunjukkan hasil transformasi dari nilai pembayaran asli (Payment Amount) setelah dinormalisasi menggunakan metode Min-Max Scaling. Berikut perintah yang dijalankan untuk proses data transformasi :

```
scaler = MinMaxScaler()
df['Payment_Scaled'] = scaler.fit_transform(df[['Payment Amount']])
print("Hasil normalisasi (nilai sudah dalam rentang 0 - 1).")
display(df.head())
```

Hasil normalisasi (nilai sudah dalam rentang 0 - 1):

	Customer	Payment Amount	Payment_Scaled
0	...	22519400.0	0.003317
1	...	35960080.0	0.003962
2	...	56265900.0	0.004937
3	...	1180833.0	0.002294
4	...	432389573.0	0.022984

Ringkasan statistik kolom hasil normalisasi:

count	7380.000000
mean	0.012156
std	0.036742
min	0.000000
25%	0.002465
50%	0.003275
75%	0.007914
max	1.000000

Name: Payment\_Scaled, dtype: float64

**Gambar 7.** Normalisasi Data

[sumber : Laseduw & Santoso, 2025]

Dari ringkasan statistik hasil normalisasi, dapat dilihat bahwa:

- Jumlah data (count) adalah 7.380 baris.
- Nilai rata-rata (mean) sebesar 0.012156, menunjukkan bahwa sebagian besar nilai pembayaran memiliki skala yang relatif kecil terhadap nilai maksimum.
- Standar deviasi (std) sebesar 0.036742, mengindikasikan adanya variasi data yang cukup tinggi antar pelanggan.
- Nilai minimum (min) adalah 0.000000 dan maksimum (max) adalah 1.000000, menandakan bahwa skala data telah berhasil diubah ke dalam rentang yang sesuai.
- Nilai kuartil (25%, 50%, 75%) memperlihatkan distribusi data yang tidak merata, di mana sebagian besar nilai berada di bawah 0.01,

menunjukkan adanya sebaran data yang cenderung condong ke bawah (*skewed*) akibat sebagian kecil pelanggan memiliki nominal pembayaran yang jauh lebih tinggi.

Secara keseluruhan, hasil ini memperlihatkan bahwa data pembayaran sudah siap untuk digunakan dalam tahap clustering, karena telah melalui proses transformasi yang memastikan skala antar nilai seragam tanpa mengubah proporsi hubungan antar data.

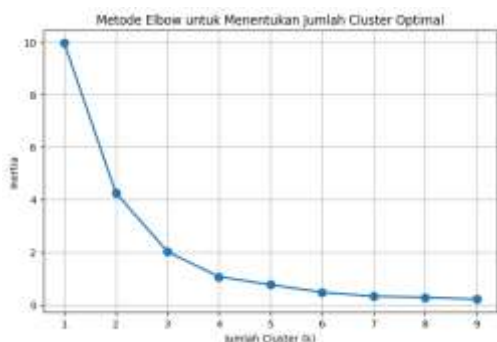
4. Data Mining

Tahap data mining merupakan inti dari proses analisis, di mana algoritma K-Means Clustering digunakan untuk mengelompokkan pelanggan berdasarkan nominal pembayaran yang telah dinormalisasi. Tujuan dari tahap ini adalah untuk mengidentifikasi pola pembayaran pelanggan dan mengelompokkan ke dalam tiga kategori utama, yaitu pelanggan dengan pembayaran rendah, sedang, dan tinggi. Sebelum menentukan jumlah cluster yang optimal, dilakukan proses pencarian nilai cluster (k) menggunakan dua metode evaluasi, yaitu :

a) *Elbow Method*

Hasil dari pencarian nilai k menggunakan *elbow method* menunjukkan bahwa dari nilai k=1 ke nilai k=3 terjadi penurunan tajam pada nilai inertia, dan setelah nilai k=3 garis mulai mendatar menunjukkan bahwa nilai inertia semakin kecil atau artinya ini menjadi indikasi bahwa nilai k=3 atau 3 cluster adalah jumlah yang optimal untuk penelitian ini. Berikut adalah perintah untuk mencari nilai k menggunakan *elbow method* :

```
inertia = []
K = range(1, 10)
for k in K:
    kmeans = KMeans(n_clusters=k,
                    random_state=42)
    kmeans.fit(df[["Payment_Scaled"]])
    inertia.append(kmeans.inertia_)
```



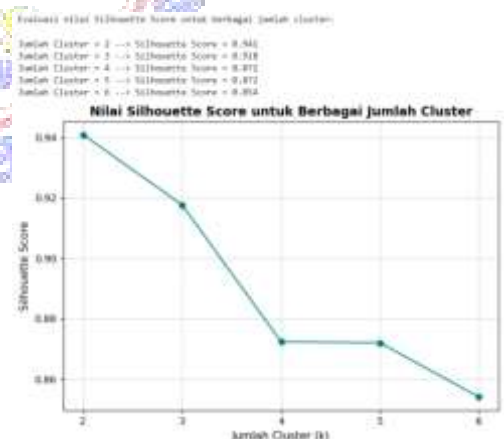
**Gambar 8.** Visualisasi Pencarian Nilai K dengan *Elbow Method*  
 [sumber : Laseduw & Santoso, 2025]

Menjelaskan pada gambar diatas bahwa sumbu X pada garis horizontal adalah jumlah cluster (k) yang dicoba, dan sumbu Y pada garis vertikal adalah nilai inertia untuk mengukur seberapa rapat data dalam setiap cluster. Pengertian inertia ialah total jarak kuadrat antar setiap data dan pusat clusternya atau centroid.

b) *Silhouette Score*

*Silhouette Score* digunakan untuk mengevaluasi kualitas pemisahan antar cluster. Hasil evaluasi menggunakan *Silhouette Score* memperlihatkan bahwa nilai untuk k = 3 merupakan yang tertinggi kedua dibandingkan jumlah cluster lainnya, sehingga masih menunjukkan kualitas pemisahan cluster yang baik. Berikut adalah perintah yang dijalankan untuk mencari *silhouette score* nya :

```
silhouette_scores = [] # simpan skor
                        # untuk tiap k
cluster_range = range(2, 7)
for k in cluster_range:
    kmeans = KMeans(n_clusters=k,
                    random_state=42)
    labels = kmeans.fit_predict(df[["Payment_Scaled"]])
    score = silhouette_score(df[["Payment_Scaled"]],
                             labels)
    silhouette_scores.append(score)
print(f"Jumlah Cluster = {k} --> Silhouette
      Score = {score:.3f}")
```



**Gambar 9.** Visualisasi *Silhouette Score* Penjelasan  
 [sumber : Laseduw & Santoso, 2025]

Penjelasan pada gambar diatas ialah sumbu X garis horizontal merupakan jumlah cluster (k) dan sumbu Y garis vertikal adalah *silhouette score*. Menunjukkan bahwa hasil dari evaluasi nilai *silhouette score* untuk tiap cluster, yaitu :

- Jumlah cluster 2 = 0.941
- Jumlah cluster 3 = 0.918
- Jumlah cluster 4 = 0.872
- Jumlah cluster 5 = 0.872
- Jumlah cluster 6 = 0.854

Semakin mendekati nilai 1, maka semakin baik hasil clustering, karena data di dalam cluster memiliki kemiripan tinggi antar anggotanya dan berbeda jauh dari cluster lain. Sebaliknya, jika nilainya mendekati 0 atau negatif, artinya pemisahan antar cluster kurang baik.

Berdasarkan kedua hasil pengujian tersebut, jumlah cluster optimal yang dipilih adalah tiga ( $k = 3$ ) karena dianggap mampu memberikan representasi yang paling sesuai untuk mengelompokkan data nominal pembayaran ke dalam tiga kategori, yaitu rendah, sedang, dan tinggi.

### 5. Interpretasi Hasil

Tahap interpretasi hasil dilakukan untuk mengimplementasikan algoritma K-Means. Setelah proses clustering dengan jumlah 3 cluster ( $k = 3$ ) dilakukan, setiap pelanggan dalam dataset mendapatkan label cluster masing-masing. Berdasarkan hasil analisis, diperoleh tiga kelompok pelanggan dengan karakteristik sebagai berikut:

- **Cluster 0** : Pelanggan dengan Pembayaran Rendah  
 Kelompok ini terdiri dari pelanggan dengan rata-rata nominal pembayaran terkecil dibandingkan cluster lainnya. Sebagian besar pelanggan dalam cluster ini melakukan pembayaran dengan nilai yang relatif kecil dan frekuensi transaksi yang lebih sering. Cluster ini menggambarkan pelanggan dengan kontribusi finansial rendah namun memiliki potensi untuk ditingkatkan melalui strategi promosi atau program loyalitas.
- **Cluster 1** : Pelanggan dengan Pembayaran Sedang  
 Kelompok ini mencerminkan pelanggan dengan nominal pembayaran menengah dan jumlah transaksi yang stabil. Pelanggan pada cluster ini merupakan segmen yang cukup penting bagi perusahaan karena kontribusi mereka signifikan dalam menjaga arus kas. Strategi yang dapat diterapkan pada kelompok ini adalah mempertahankan loyalitas pelanggan melalui pelayanan yang konsisten dan program insentif.

- **Cluster 2** : Pelanggan dengan Pembayaran Tinggi  
 Kelompok ini terdiri dari pelanggan dengan nominal pembayaran tertinggi. Meskipun jumlahnya relatif lebih sedikit, pelanggan pada cluster ini memberikan kontribusi besar terhadap total pendapatan perusahaan. Kelompok ini dapat dianggap sebagai pelanggan prioritas yang memerlukan pendekatan layanan khusus, seperti personalisasi layanan atau penawaran eksklusif, untuk menjaga hubungan jangka panjang.

Berdasarkan hasil proses klasterisasi menggunakan algoritma K-Means, data pelanggan berhasil dikelompokkan ke dalam beberapa kategori 3 klaster yang telah dijelaskan.

Customer	Payment Amount	Payment_Scaled	Cluster	Kategori
0	32519400.0	0.003217	0	Cluster 0 (Rendah)
1	35960080.0	0.003962	0	Cluster 0 (Rendah)
2	36283800.0	0.004027	0	Cluster 0 (Rendah)
3	1180033.0	0.000294	0	Cluster 0 (Rendah)
4	43229573.0	0.022884	0	Cluster 0 (Rendah)

**Gambar 10.** Contoh Hasil Klasterisasi pada Data Head

[sumber : Laseduw & Santoso, 2025]

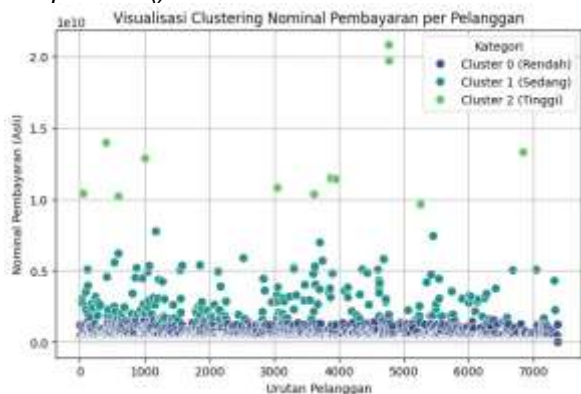
Hasil di atas menunjukkan contoh klasterisasi pada lima data pertama. Hal ini menunjukkan bahwa algoritma K-Means mampu mengelompokkan pelanggan dengan karakteristik nominal pembayaran yang serupa ke dalam satu klaster yang sama.

#### a) Visualisasi

Visualisasi hasil proses clustering dilakukan dengan menggunakan scatter plot dan box plot. Scatter plot digunakan untuk menggambarkan sebaran data pelanggan berdasarkan nominal pembayaran yang telah dikelompokkan ke dalam tiga cluster, sehingga pola dan perbedaan antar kategori dapat diamati secara visual. Berikut perintah untuk menjalankan visualisasi clustering dengan scatter plot:

```
kmeans = KMeans(n_clusters=3,
random_state=42)
df['Cluster'] = kmeans.fit_predict(df[['Payment_Scaled']])
mapping = {0: 'Cluster 0 (Rendah)', 1: 'Cluster 1 (Sedang)', 2: 'Cluster 2 (Tinggi)'}
df['Kategori'] = df['Cluster'].map(mapping)
df_clustered = df.copy()
df_clustered['Customer_Index'] = range(len(df_clustered))
plt.figure(figsize=(8,5))
sns.scatterplot
```

```
x='Customer_Index',
y='Payment Amount',
hue='Kategori',
palette='viridis',
data=df_clustered,
s=60)
plt.title("Visualisasi Clustering Nominal
Pembayaran per Pelanggan")
plt.xlabel("Urutan Pelanggan")
plt.ylabel("Nominal Pembayaran (Asli)")
plt.legend(title="Kategori")
plt.grid(True)
plt.show()
```



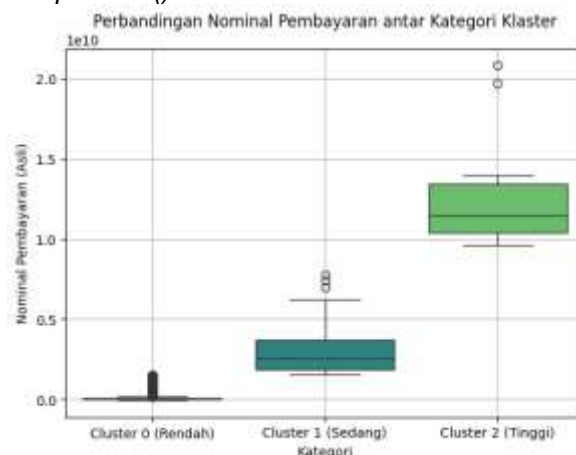
**Gambar 11.** Visualisasi Scatter Plot  
 [sumber : Laseduw & Santoso, 2025]

Berdasarkan hasil visualisasi clustering pada gambar diatas terlihat bahwa proses pengelompokan data nominal pembayaran dengan algoritma K-Means menghasilkan tiga kategori utama, yaitu cluster 0 (rendah), cluster 1 (sedang), dan cluster 2 (tinggi). Sebagian besar pelanggan termasuk dalam kategori rendah yang ditunjukkan oleh sebaran titik berwarna ungu di bagian bawah grafik, sedangkan kategori sedang berwarna hijau tua dan tinggi ditunjukkan oleh titik berwarna hijau muda yang posisinya lebih menyebar pada nilai nominal yang lebih besar. Distribusi ini mengindikasikan bahwa transaksi dengan nominal kecil mendominasi data, namun pelanggan dengan pembayaran tinggi memiliki nilai kontribusi yang lebih signifikan terhadap total pembayaran perusahaan.

Sementara itu, box plot digunakan untuk menampilkan distribusi nilai pembayaran pada setiap cluster, yang mencakup nilai median, rentang antar kuartil, serta identifikasi nilai ekstrem (outlier). Perintah yang digunakan untuk menjalankan visualisasi dengan box plot, ialah:

```
plt.figure(figsize=(7,5))
sns.boxplot(
x='Kategori',
y='Payment Amount',
data=df_clustered,
palette='viridis')
plt.title("Perbandingan Nominal
Pembayaran antar Kategori Klaster")
plt.xlabel("Kategori")
```

```
plt.ylabel("Nominal Pembayaran (Asli)")
plt.grid(True)
plt.show()
```



**Gambar 12.** Visualisasi Box Plot  
 [sumber : Laseduw & Santoso, 2025]

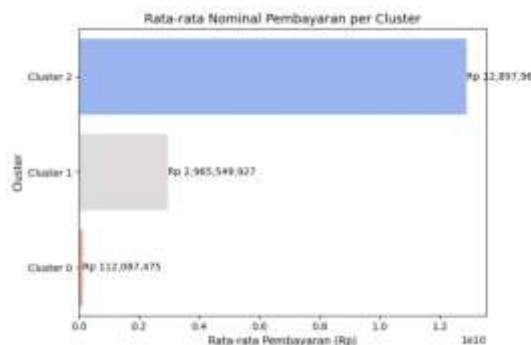
Visualisasi menggunakan boxplot pada Gambar diatas menunjukkan perbandingan distribusi nominal pembayaran pada setiap kategori klaster. Hasilnya memperlihatkan bahwa Cluster 0 (Rendah) memiliki nilai median dan sebaran yang kecil, menandakan kelompok pelanggan dengan nominal pembayaran rendah dan relatif homogen. Cluster 1 (Sedang) memiliki median yang lebih tinggi serta rentang data yang lebih luas, menggambarkan variasi nominal pembayaran yang cukup besar pada kelompok pelanggan menengah. Sementara itu, Cluster 2 (Tinggi) menampilkan median dan sebaran nilai pembayaran tertinggi, dengan beberapa data ekstrem (outlier) yang menunjukkan adanya pelanggan dengan nominal pembayaran sangat besar.

Secara keseluruhan, hasil visualisasi ini memperkuat temuan dari proses clustering, kedua visualisasi ini saling melengkapi dalam memberikan pemahaman terhadap karakteristik masing-masing kelompok pelanggan hasil dari penerapan algoritma K-Means.

b) Rata-Rata Pembayaran

Penggunaan nilai rata-rata nominal pembayaran pada setiap cluster bertujuan untuk menggambarkan kecenderungan umum dari kelompok pelanggan yang terbentuk. Rata-rata ini menjadi indikator utama dalam menilai perbedaan karakteristik antar cluster, serta membantu perusahaan dalam mengidentifikasi kelompok pelanggan berdasarkan tingkat kontribusi pembayarannya. Selain itu, nilai rata-rata yang berbeda secara signifikan memperkuat validitas hasil klasterisasi yang dilakukan dengan algoritma K-Means. Berikut perintah yang dijalankan untuk mengetahui rata-rata pembayaran pada setiap cluster :

```
cluster_means = (
df.groupby('Cluster')['Payment Amount']
.mean()
.sort_values(ascending=False)
print("\nRata-rata nominal pembayaran per cluster
(dalam Rupiah):")
for cluster, nilai in cluster_means.items():
print(f"Cluster {cluster}: Rp {nilai:,.0f}")
sns.barplot(
x=cluster_means.values,
y=[f"Cluster {i}" for i in cluster_means.index],
palette="coolwarm"
for i, value in enumerate(cluster_means.values):
plt.text(value, i, f"Rp {value:,.0f}", va='center',
ha='left', fontsize=10)
plt.title("Rata-rata Nominal Pembayaran per
Cluster", fontsize=13)
plt.xlabel("Rata-rata Pembayaran (Rp)",
fontsize=11)
plt.ylabel("Cluster", fontsize=11)
```



**Gambar 13.** Visualisasi Rata-Rata Pembayaran per Cluster  
 [sumber : Laseduw & Santoso, 2025]

Berdasarkan gambar diatas, dapat dilihat bahwa nilai rata-rata nominal pembayaran berbeda secara signifikan pada setiap cluster. Cluster 0 (Rendah) memiliki rata-rata nominal pembayaran sebesar Rp112.087.475, yang menunjukkan kelompok pelanggan dengan tingkat pembayaran paling rendah. Cluster 1 (Sedang) memiliki rata-rata sebesar Rp2.965.549.927, menggambarkan pelanggan dengan aktivitas pembayaran menengah. Sementara itu, Cluster 2 (Tinggi) memiliki rata-rata tertinggi yaitu sebesar Rp12.897.962.436, yang merepresentasikan kelompok pelanggan dengan kontribusi pembayaran terbesar terhadap total transaksi.

**KESIMPULAN**

Penelitian ini berfokus pada implementasi algoritma Kmeans untuk clustering nominal pembayaran pelanggan pada PT. XYZ. Dapat disimpulkan bahwa penerapan algoritma K-Means mampu memberikan hasil pengelompokan (clustering) yang efektif terhadap data nominal pembayaran pelanggan. Proses pengolahan data dilakukan menggunakan pendekatan KDD (Knowledge Discovery in Database) yaitu melalui tahap data seleksi data, preprocessing data, transformasi data, data mining atau implementasi

K-Means dan interpretasi hasil. Hasil analisis menggunakan Elbow Method dan Silhouette Score menunjukkan bahwa jumlah kluster optimal adalah tiga (k=3), yang merepresentasikan kategori pembayaran rendah, sedang, dan tinggi. Implementasi hasil klustering ini memberikan nilai tambah bagi perusahaan dalam hal analisis dan pengambilan keputusan. Pihak manajemen dapat dengan mudah memantau pola pembayaran pelanggan dan menetapkan strategi bisnis yang lebih tepat sasaran. Cluster dengan nominal tinggi dapat dijadikan prioritas dalam pengelolaan relasi pelanggan, sedangkan kluster dengan nominal rendah dapat menjadi fokus evaluasi dan pengembangan strategi peningkatan penjualan.

**DAFTAR PUSTAKA**

[1] A. Aditya and H. F. Efendi, "Business Process Analysis and Implementation of Odoo Open Source ERP System in Inventory, Purchasing and Sales Activities," *Procedia Soc. Sci. Humanit.*, vol. 3, no. January, pp. 349–357, 2022, doi: 10.21070/pssh.v3i.180.

[2] Y. Eko, H. Santoso, and A. Aditya, "Procedia of Social Sciences and Humanities Business Process Analysis and Implementation of Odoo ERP in Sales , Purchasing and Accounting Systems ( Case Study : CV Mitra Perkasa ) Analisis Proses Bisnis dan Penerapan ERP Odoo pada Sistem Penjualan , Pembelian dan Akuntansi Studi Kasus CV . Mitra Procedia of Social Sciences and Humanities," vol. 0672, no. c, pp. 358–365, 2022.

[3] Maulana Ardiansyah and Supatman, "KLAUSTERISASI, TAGIHAN PADA NASABAH PINJAMAN ONLINE MENGGUNAKAN METODE K-MEANS CLUSTERING," *J. RESTIKOM Ris. Tek. Inform. dan Komput.*, vol. 6, no. 2, pp. 286–294, 2024, doi: 10.52005/restikom.v6i2.318.

[4] H. Mutaqin, "Penerapan Algoritma K-Means Pada Penjualan Saldo Transportasi Online Studi Kasus Konter XYZ," *J. Algoritm.*, vol. 20, no. 1, pp. 57–64, 2023, doi: 10.33364/algoritma/v.20-1.1244.

[5] S. W. Harjono, N. W. Utami, and I. G. A. P. D. Putri, "Klasterisasi Tingkat Penjualan pada Startup Panak.id dengan Algoritma K-Means," *J. Ilm. Teknol. Inf. Asia*, vol. 17, no. 1, pp. 55–66, 2023, doi: 10.32815/jitika.v17i1.888.

[6] A. A. Alolayan and A. A. Alhamed, "Detection of Knowledge on Social Media Using Data Mining Techniques," pp. 472–482, 2024, doi: 10.4236/ojapps.2024.142034.

[7] T. Gori et al., "PREPROCESSING DATA DAN KLASIFIKASI UNTUK PREDIKSI KINERJA DATA PREPROCESSING AND CLASSIFICATION FOR PREDICTING

- STUDENT,” vol. 11, no. 1, pp. 215–224, 2024, doi: 10.25126/jtiik.20241118074.
- [8] R. Sistem, “Jurnal resti,” vol. 1, no. 10, pp. 117–122, 2021.
- [9] G. G. Dodi Alexsander Manalu, “IMPLEMENTASI METODE DATA MINING K-MEANS CLUSTERING TERHADAP DATA PEMBAYARAN TRANSAKSI MENGGUNAKAN BAHASA PEMROGRAMAN PYTHON PADA CV DIGITAL DIMENSI,” *Raspberry Pi OS Syst. Adm. with Syst. Python*, vol. 8, no. 1, pp. 175–305, 2023, doi: 10.1201/b23421-3.
- [10] M. R. Syahkur and D. Hartama, “Evaluasi Jumlah Cluster pada Algoritma K-Means ++ Menggunakan Silhouette dan Elbow dengan Validasi Nilai DBI dalam Mengelompokkan Gizi Balita,” vol. 13, no. 3, pp. 487–496, 2024.

