

# Pengaruh Outlier pada Algoritma K-Medoid untuk Mengelompokkan Rekanan Vendor dalam Pengadaan Barang

1<sup>st</sup> Nopiti Yulistiani  
Universitas Buana Perjuangan  
Karawang, Indonesia  
if.nopitiyulistiani@mhs.ubpkarawang.ac.id

2<sup>rd</sup> Ayu Ratna Juwita  
Universitas Buana Perjuangan  
Karawang, Indonesia  
ayurj@ubpkarawang.ac.id

3<sup>rd</sup> Anis Fitri Nur Masruriyah  
Universitas Buana Perjuangan  
Karawang, Indonesia  
anis.masruriyah@ubpkarawang.ac.id

**Abstract**— Dalam dunia bisnis, vendor memainkan peran penting dalam kinerja perusahaan. Pemilihan vendor yang tepat sangat penting untuk memastikan kelancaran pelaksanaan proyek dalam pengadaan barang. Jika vendor dipilih secara sembarangan, maka perusahaan dapat mengalami kerugian dalam berbagai aspek seperti kualitas, biaya, dan waktu. Oleh karena itu, diperlukan metode yang tepat untuk mengelompokkan vendor agar dapat mengatasi masalah tersebut. Salah satu teknik yang dapat digunakan adalah clustering untuk mengelompokkan data vendor menjadi tiga kelompok, yaitu tinggi, sedang, dan rendah. Metode clustering yang digunakan dalam penelitian ini adalah algoritma k-medoids. Hasil penelitian menunjukkan bahwa kualitas pengelompokan vendor menggunakan teknik penghapusan outliers lebih baik daripada menggunakan data dengan outliers. Kontribusi utama dari penelitian ini adalah untuk memverifikasi apakah kualitas cluster tanpa outlier lebih tinggi dibandingkan dengan outlier dalam data dengan membandingkan dan menganalisis algoritma k-medoids dengan dan tanpa deteksi outliers tergantung pada algoritma pengelompokan yang diterapkan dan data yang digunakan. Selanjutnya, hasil penelitian menunjukkan bahwa algoritma k-medoids menghasilkan nilai Silhouette Coefficient sebesar 0,6454.

**Kata kunci** — Clustering, K-medoids, Outliers, Silhouette Coefficient.

## I. PENDAHULUAN

Peran vendor memiliki bagian penting dalam memastikan kelancaran proses produksi, karena bahan mentah selalu diperlukan sebagai bahan baku dalam proses produksi [1]. Agar mencapai hasil produksi yang diinginkan, penting untuk memastikan ketersediaan bahan baku yang diperlukan selama proses produksi. Oleh karena itu, pemilihan vendor yang tepat menjadi suatu keharusan agar kebutuhan bahan baku tersebut dapat terpenuhi [2]. Agar mendapatkan vendor yang sesuai, perlu mempertimbangkan kriteria yang relevan dengan kebutuhan perusahaan. Beberapa kriteria yang dapat dipertimbangkan dalam proses pemilihan vendor yaitu meliputi kualitas produk, layanan, reputasi, pengiriman tepat waktu, serta harga yang bersaing [3]. Namun, perusahaan yang bergerak di bidang percetakan sering mengalami kesulitan dalam menentukan vendor yang akan diundang untuk memasok barang pada saat melakukan pengadaan. Hal ini disebabkan oleh banyaknya jumlah vendor yang telah terdaftar di perusahaan, sehingga pihak pengadaan kesulitan untuk menentukan dan memilih rekanan vendor yang tepat [4]. Jika perusahaan mengundang vendor yang tidak responsif dan tidak konsisten dalam bekerja, hal tersebut dapat menghambat kinerja perusahaan. Oleh karena itu, perlu dilakukan pengelompokan vendor berdasarkan vendor potensial tinggi, sedang, dan rendah.

Penelitian sebelumnya oleh Rhamadani et al [5] menjelaskan perbandingan algoritma k-means dan k-medoids untuk mengelompokkan nilai ujian nasional, Peneliti melakukan pengelompokan sekolah menjadi tiga kategori (baik, sedang, dan cukup) menggunakan algoritma K-Means dan K-Medoid, data hasil Ujian Nasional yang disajikan oleh Pusat Penilaian Pendidikan tidak memiliki peringkat atau kelompok berdasarkan kategori nilai UN. Hasil pengelompokan menunjukkan perbedaan jumlah kluster pada kinerja tiap algoritma memiliki pola perhitungan yang berbeda tergantung pada dataset yang digunakan dan titik centroid pada algoritma. Selanjutnya agneresa et al [6] melakukan strategi promosi pada mahasiswa baru dengan menggunakan algoritma k-means pada penelitian ini menggunakan rapidminer untuk implementasi. Davies Bouldin Index digunakan untuk mencari jumlah Cluster yang optimal kemudian menghasilkan 2 Cluster, Cluster 1 sebagai cluster minat tinggi berjumlah 1945 data, Cluster 2 sebagai cluster minat rendah berjumlah 525 data dengan presentase masing-masing cluster 1 78,82% dan cluster 2 21,18 %. Penelitian ini menghasilkan pola yang menarik sebagai sumber informasi baru untuk menentukan strategi promosi yang lebih efektif dan tepat sasaran dalam mempromosikan setiap program studi sehingga dapat meningkatkan jumlah calon pendaftar mahasiswa baru di tahun berikutnya. Kemudian Rahmawati et al [7] melakukan pemilihan rekomendasi vendor online. Penelitian tersebut menggunakan algoritma Decision Tree, dengan perhitungan Entropy. Nilai akurasi dihitung menggunakan perhitungan Cross Validation, Pruning dan Prepruning. Hasil perhitungan dari teknik Pruning dan Prepruning memiliki tingkat error yang lebih rendah dengan selisih nilai 1.34%. Sedangkan teknik Cross Validation memiliki tingkat akurasi 91.51%. Selanjutnya Ummah et al [8] dalam penelitian ini, dilakukan rekomendasi vendor proyek menggunakan algoritma Naïve Bayes. Menentukan tingkat akurasi, penelitian ini menghitung nilai Probabilitas menggunakan metode Naïve Bayes dan menggunakan confusion matrix. Hasil penelitian menunjukkan bahwa nilai Probabilitas class Terima 0,725925 dan class Tidak Terima 0,274075. Tingkat akurasi yang dihasilkan menggunakan confusion matrix sebesar 86,67%.

Berdasarkan permasalahan dalam pemilihan rekanan vendor dan solusi dari penelitian sebelumnya, penelitian ini bertujuan untuk mengelompokkan rekanan vendor berdasarkan evaluasi kinerja. Berbeda dengan penelitian sebelumnya, penelitian ini memiliki fokus pada pengelompokan rekanan vendor yang dapat membantu pihak pengadaan dalam menentukan rekanan yang akan diundang pada pengadaan selanjutnya.

## II. DATA DAN MODE

### A. Vendor

Vendor merupakan bagian penting dari rantai pasokan yang dapat mempengaruhi kinerja perusahaan [9]. Vendor merupakan pemasok barang atau jasa yang menjalin kontrak kerja dengan perusahaan [1]. Perusahaan menetapkan persyaratan yang harus dipenuhi oleh calon vendor sesuai standar kerja. Kemudian, calon vendor memperkenalkan produk dan jasa yang ditawarkan, disertai dengan profil perusahaan dan dokumen legal yang lengkap. [10] Pemilihan vendor merupakan kegiatan penting bagi perusahaan untuk memastikan kelancaran proyek dan penyelesaian tepat waktu. Produk berkualitas tinggi yang sesuai dengan permintaan pelanggan dapat menjamin reputasi perusahaan yang baik dan mempertahankan penilaian yang baik.

### B. Data Mining

Data mining adalah proses analisis dari sekumpulan data yang bertujuan untuk menemukan hubungan yang tidak terduga dan meringkas data dengan cara yang berbeda. Hal ini dapat dipahami dan bermanfaat bagi pemilik data [1]. Proses data mining bersifat iterative dan interaktif, dimana tujuannya adalah untuk menemukan pola atau model baru yang sempurna [11]. Data mining sangat bermanfaat dalam database yang sangat besar (massive database), dimana pencarian trend atau pola yang diinginkan dapat membantu pengambilan keputusan di masa depan. Pola-pola tersebut dapat dikenali oleh perangkat tertentu dan memberikan analisa data yang berguna dan berwawasan [11].

### C. Algoritma K-medoids

Algoritma PAM adalah salah satu metode clustering partitioning yang digunakan untuk mengelompokkan suatu kumpulan objek sebanyak n menjadi k cluster. Pada algoritma ini, representasi cluster dilakukan dengan memilih salah satu objek sebagai medoid yang merepresentasikan objek-objek pada cluster tersebut [12]. K-medoids mempunyai sifat bahwa pusat cluster berada di tengah-tengah titik data. Algoritma ini menggunakan jarak Euclidean untuk menghitung kedekatan objek dengan pusat (medoid) sehingga objek yang dipilih secara acak cenderung mirip dengan objek medoid. Oleh karena itu, algoritma k-medoids memiliki kelebihan dibandingkan dengan algoritma k-means [13]. Gambar 1 menunjukkan *pseudocode* untuk langkah-langkah menyelesaikan *K-medoids*.

---

**ALGORITHM 2:** item clustering algorithm

**Input:** user-item rating matrix  $R$  ( $M$  users and  $N$  items),  $k$  cluster centers  $C$  and KL distance measure.

**Output:**  $k$  clustering results

**Time complexity:**  $O(N)$

```

1 for  $j \leftarrow 1$  to  $N$  do
2   if  $j$  is not equal to any of cluster centers in  $C$  then
3     Set  $t=1$ ;
4     Set  $D[t]=D(j|C[t])$ ;
5     for  $n \leftarrow 2$  to  $k$  do
6        $D[n]=D(j|C[n])$ ;
7       if  $D[t] \geq D[n]$  then
8          $t=n$ ;
9       end
10    end
11     $j \in C_t$  // item  $j$  is assigned to the nearest cluster center  $C_t$ 
12  end
14 end
    
```

---

Gambar 1 Pseudocode K-Medoids

### D. Evaluasi Cluster

evaluasi clustering akan dilakukan menggunakan Silhouette Coefficient. Setelah mendapatkan hasil perhitungan data mining yang sesuai, akan dilakukan evaluasi menggunakan Silhouette Coefficient untuk menilai validitas algoritma dalam mengelompokkan cluster. Silhouette Coefficient memiliki rentang nilai antara -1 hingga 1, dimana semakin mendekati 1 menunjukkan bahwa pengelompokan data dalam satu cluster semakin baik. Sebaliknya, semakin mendekati -1 menunjukkan bahwa pengelompokan data dalam satu cluster semakin buruk. Langkah-langkah perhitungan yang digunakan dalam *Silhouette coefficient* adalah sebagai berikut [14] :

1. Menghitung rata-rata jarak dari suatu dokumen misalkan  $i$  dengan semua dokumen lain yang berada dalam satu *cluster*, dengan persamaan (1).

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j) \tag{1}$$

dengan  $j$  adalah dokumen lain dalam satu *cluster*  $A$  dan  $d(i,j)$  adalah jarak antara dokumen  $i$  dengan  $j$ .

2. Menghitung rata-rata jarak dari dokumen  $i$  tersebut dengan semua dokumen di *cluster* lain, dan diambil nilai terkecilnya, menggunakan persamaan (2).

$$d(i, C) = \frac{1}{|A|} \sum_{j \in C} d(i, j) \tag{2}$$

dengan  $d(i,C)$  adalah jarak rata-rata dokumen  $i$  dengan semua objek pada *cluster* lain  $C$  dimana  $A \neq C$ , dapat dihitung menggunakan persamaan (3).

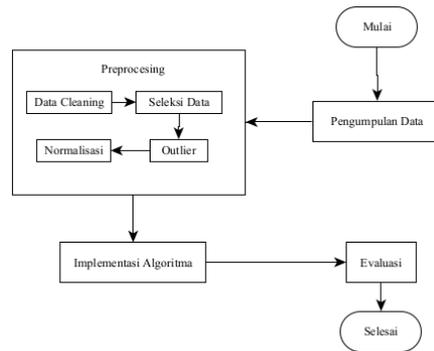
$$b(i) = \min_{C \neq A} d(i, C) \tag{3}$$

3. Nilai *Silhouette Coefficient* menghitung menggunakan persamaan (4) :

$$s(i) = \frac{d(i) - a(i)}{\max(a(i), b(i))} \tag{4}$$

E. *Prosedur Penelitian*

Prosedur penelitian dimulai dengan studi literatur dengan rincian penelitian sebagai berikut:



Gambar 2 *Prosedur Penelitian*

III. HASIL DAN PEMBAHASAN

A. *Pengumpulan data*

Data diperoleh dari perusahaan yang bergerak dibidang percetakan sebagai acuan data dalam penelitian ini dan data diproses menjadi beberapa proses. Data yang diperoleh terdapat 9 kolom (SAP, vendor, nilai awal, prestasi, pelanggaran, total nilai, spph menawarkan, dan spph ditawarkan) terdiri dari 770 data.

Tabel 1 *Dataset*

No	SAP	VENDOR	AWAL	PRESTASI	PELANGGARAN	TOTAL NILAI	SPPH DITAWARKAN	SPPH MENAWARKAN
1	0010000202	PT.A	75	1	0	78	95	100
2	0010001009	PT.AAS	75	1	0	78	11	454
3	0010000159	PT.ADH	75	0	0	75	77	644
4	0010000248	PT.ADT	75	4	0	80	202	622
5	0010001241	PT.ARS	75	0	0	75	126	574
6	0010000115	PT.AI	75	0	0	75	27	614
7	0010000782	PT.A	75	5	0	80	28	720
8	0010001200	PT.AK	75	0	0	75	47	441
9	0010000931	PT.AT	75	0	0	75	48	88
...								
768	0010001049	PT.ZKS	75	0	0	75	4	3
769	0010000077	PT.ZGA	75	0	-3	72	4	3
770	0010000087	PT.YWI	75	0	-3	72	4	3

B. *Data Cleaning*

Pada langkah data *cleaning*, dilakukan upaya untuk menghapus data yang tidak relevan seperti data yang hilang, *noise* data, dan data yang tidak konsisten. dilakukan pemeriksaan nilai-nilai yang hilang pada dataset untuk memastikan tidak ada data yang hilang atau kosong. Berikut hasil check missing value terdapat pada Gambar 3.

```
No. 0
SAP 2
VENDOR 0
VENDOR.1 0
AWAL 0
PRESTASI 0
PELANGGARAN 0
TOTAL NILAI 0
SPPHDITAWARKAN 0
MENAWARKAN 0
KETERANGAN 0
dtype: int64
```

Gambar 3 Cek Missing value

Dataset yang telah digunakan dalam penelitian memiliki nilai yang hilang atau terdapat missing values. Tabel 2 merupakan data yang nilainya kosong.

Tabel 2 Missing Value

No	SAP	VENDOR	AWAL	PRESTASI	PELANGGARAN	TOTAL NILAI	SPPHDITAWARKAN	SPPHMENAWARKAN
290	?	IP	75	0	0	75	0	1
605	?	RF	75	0	0	75	8	18

Keterangan hasil untuk mengatasi missing value, dengan cara mengambil rata-rata dari setiap attribute yang ada, dapat dilihat pada Tabel 3.

Tabel 3 Mengatasi Missing Value

No	SAP	VENDOR	AWAL	PRESTASI	PELANGGARAN	TOTAL NILAI	SPPHDITAWARKAN	SPPHMENAWARKAN
290	10000608.553385416	IP	75	0	0	75	0	1
605	10000608.553385416	RF	75	0	0	75	8	18

C. Seleksi Data

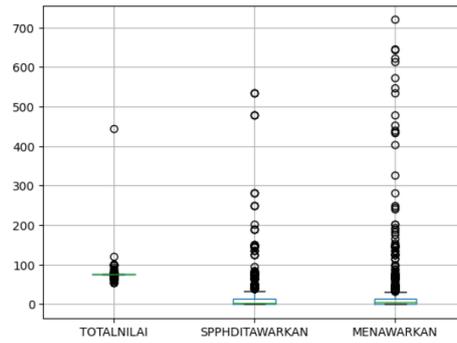
Langkah berikutnya setelah proses data cleaning yaitu seleksi data, di mana data akan dipilih dengan cermat dari dataset, karena tidak semua data diperlukan. Proses seleksi data ini sangat penting karena akan mempengaruhi hasil analisis dari penelitian yang dilakukan, dapat dilihat pada tabel 4.

Tabel 4 Hasil Seleksi Data

TOTAL NILAI	SPPH DITAWARKAN	SPPHMENAWARKAN
78	95	100
78	11	454
75	77	644
80	202	622
75	126	574
75	27	614
80	28	720
75	47	441
75	48	88
.	.	.
.	.	.
75	4	3
72	4	3
72	4	3

D. Outliers

Data outliers merujuk pada data yang memiliki nilai yang sangat berbeda dibandingkan dengan sebagian besar data yang ada. Salah satu teknik yang dapat digunakan untuk menangani outliers dengan melakukan penyaringan menggunakan quantile. Gambar 4 hasil dari cek outliers dari setiap variabel. Akan diatasi dengan teknik quartile agar data tidak lagi berjarak jauh dengan data lainnya. Teknik yang digunakan untuk melakukan filter quantile dengan Inter Quartile Range (IQR) dengan bantuan visualisasi data dalam bentuk boxplot. IQR dihitung dengan mengurangkan kuartil ketiga (Q3) dari kuartil pertama (Q1).



Gambar 4 cek outliers

Dari ilustrasi yang diberikan untuk ketiga variabel, terlihat bahwa terdapat banyak data yang tergolong sebagai outliers, yang dimana outlier ini data yang menyimpang secara ekstrim dari rata rata sekumpulan data yang ada. Oleh karena itu, ketika dilakukan perbandingan hasil menggunakan teknik yang mengabaikan outliers dan menggunakan teknik yang memperhitungkan outliers, akan terdapat dua hasil yang berbeda. Perbandingan dapat dilihat pada langkah berikutnya.

E. Normalisasi

penelitian ini, dilakukan normalisasi data untuk mengubah nilai variabel menjadi skala umum dengan menggunakan teknik min-max, dengan tujuan agar rentang nilai variabel berada dalam kisaran 0 dan 1. Pada proses normalisasi, akan dilakukan perbandingan hasil antara data yang menggunakan teknik outliers dan data yang tidak menggunakan teknik outliers. Perbandingan tersebut menunjukkan perbedaan nilai yang sangat berbeda antara kedua kelompok data. Hasil yang tidak menggunakan teknik outliers dapat ditemukan di Tabel 5, sementara hasil yang menggunakan teknik outliers dapat ditemukan di Tabel 6. Kedua tabel tersebut menunjukkan perbedaan yang signifikan.

Tabel 5 Normalisasi sebelum pakai outliers

TOTALNILAI	SPPH DITAWARKAN	SPPH MENAWARKAN
0.061538	0.177570	0.138889
0.061538	0.020561	0.630556
0.053846	0.143925	0.894444
0.066667	0.377570	0.863889
0.053846	0.235514	0.797222
...	...	...
0.058974	0.007477	0.004167
0.053846	0.007477	0.004167
0.053846	0.007477	0.004167
0.046154	0.007477	0.004167
0.046154	0.007477	0.004167

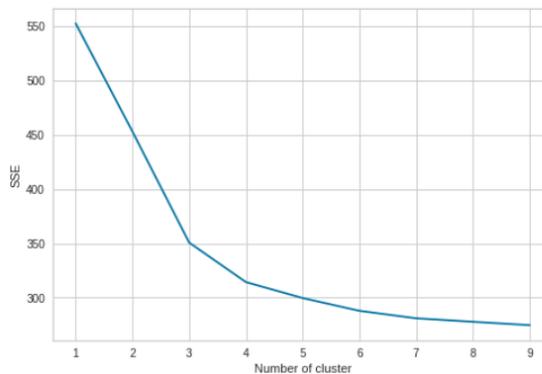
Tabel 6 Normalisasi pakai Teknik outliers

TOTALNILAI	SPPH DITAWARKAN	SPPH MENAWARKAN
0.061538	0.177570	0.138889
0.061538	0.020561	0.630556
0.053846	0.143925	0.894444
0.066667	0.377570	0.863889
0.053846	0.235514	0.797222
...	...	...
0.058974	0.007477	0.004167
0.053846	0.007477	0.004167
0.053846	0.007477	0.004167
0.046154	0.007477	0.004167
0.046154	0.007477	0.004167

Setelah melalui tahap preprocessing tersebut, langkah selanjutnya tahapan data mining menggunakan algoritma K-Medoids

F. Algoritma K-Medoids

Penentuan jumlah cluster pada penelitian ini menggunakan metode Elbow, Proses terakhir melihat hasil plot yang dihasilkan. Pencarian nilai k optimum dilakukan dengan membandingkan nilai SSE (Sum of Square Error) yang disajikan dalam bentuk grafik. Pada tahapan ini dilakukan pengelompokan data evaluasi vendor dengan melakukan pemodelan percobaan dari K1 hingga K9.



Gambar 5 Hasil Metode Ellbow

Setelah itu, program akan menampilkan output berupa nilai centroid terakhir dalam dua bentuk, yaitu Tabel 7 yang menunjukkan hasil centroid tanpa menggunakan teknik outliers dan Tabel 8 yang menunjukkan hasil centroid dengan menggunakan teknik outliers. Terdapat perubahan nilai centroid yang signifikan antara kedua tabel tersebut pada algoritma k-medoids.

Tabel 7 Nilai Centroid Sebelum memakai Outliers

array	$\left( \begin{aligned} &[[-0.03689168, 0.92619946, -0.29237583], \\ &[-0.03689168, -0.32834484, -0.26725539], \\ &[-0.03689168, -0.25864793, 2.10662623]] \end{aligned} \right)$
-------	---

Tabel 8 Nilai Centroid Sesudah memakai Outliers

array	$\left( \begin{aligned} &[[ 0. , -0.13421755, 1.62934695], \\ &[ 0. , -0.63068278, -0.61887388], \\ &[ 0. , 1.93438757, -0.61887388]] \end{aligned} \right)$
-------	--

Kemudian terdapat jumlah cluster yang memiliki dua perbedaan, tabel 9 hasil dari jumlah cluster yang tidak menggunakan teknik outliers, tabel 10 jumlah cluster yang memakai teknik outliers. jumlah kedua tabel tersebut menghasilkan jumlah yang sangat berbeda pada pengelompokan rekanan vendor.

Tabel 9 Jumlah Sebelum Memakai Outliers

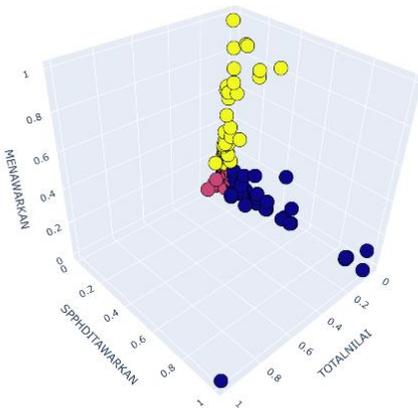
Jumlah Anggota Setiap Cluster	
Rekanan Potensial Tinggi	103
Rekanan Potensial Sedang	625
Rekanan Potensial Rendah	42

Tabel 10 Tabel 11 Jumlah Sesudah Memakai Outliers

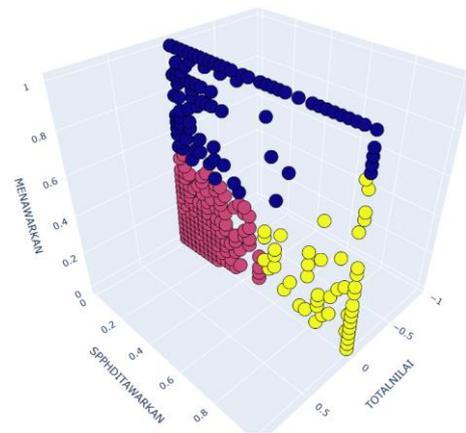
Jumlah Anggota Setiap Cluster	
Rekanan Potensial Tinggi	178
Rekanan Potensial Sedang	473
Rekanan Potensial Rendah	119

Dari tabel 9 dan 10, terlihat bahwa pengelompokan cluster tanpa menggunakan teknik outliers menghasilkan 103 rekanan potensial tinggi, 625 rekanan potensial sedang, dan 42 rekanan potensial rendah. Namun, pengelompokan tanpa teknik outliers ini menghambat pembentukan cluster. Oleh karena itu, masalah outliers dalam data sangat penting karena outlier dapat memengaruhi kesulitan dalam membuat kelompok. Sementara itu, pada hasil pengelompokan yang menggunakan teknik penghilangan outliers, terlihat bahwa terdapat 178 rekanan potensial tinggi, 473 rekanan potensial sedang, dan 119 rekanan potensial rendah. Setelah dilakukan penghapusan outlier, cluster menjadi lebih konsisten, dengan kata lain kesamaan objek dalam kelompok menjadi lebih signifikan daripada sebelum dilakukan penghapusan outlier.

Selanjutnya, terdapat dua visualisasi yang menunjukkan perbedaan hasil dari pengelompokan dengan menggunakan algoritma k-medoids. Visualisasi pertama pada Gambar 6 yang menunjukkan hasil pengelompokan tanpa menggunakan teknik outliers, sedangkan visualisasi kedua pada Gambar 7 yang menunjukkan hasil pengelompokan dengan menggunakan teknik outliers. Perbedaan antara kedua visualisasi ini sangat berbeda dalam pengelompokan.



Gambar 6 Visualisasi Sebelum Memakai Outlier



Gambar 7 Gambar 8 Visualisasi Sesudah Memakai Outlier

Pada gambar 6 dan 7, terdapat beberapa perbedaan. Gambar 6 menunjukkan banyak adanya outliers yang menyebabkan kesulitan dalam pembuatan cluster. Oleh karena itu, masalah outliers dalam data sangat berpengaruh dalam kualitas cluster. Outlier dapat berupa kesalahan input data atau noise informasi dan outlier data sebenarnya. Kualitas cluster yang mengandung outliers dapat dibandingkan dengan kualitas cluster setelah dilakukan penghilangan outliers. Sementara itu, pada gambar 7 terlihat bahwa teknik penghilangan outliers telah digunakan. Cluster setelah dilakukan penghilangan outliers memiliki kualitas yang lebih baik karena lebih konsisten secara internal dan dipisahkan dengan baik secara eksternal. Dengan demikian, pencarian dalam cluster menjadi lebih efisien.

### G. Evaluasi

Setelah dilakukan evaluasi kluster menggunakan metode koefisien siluet, diperoleh nilai koefisien siluet sebagai berikut. Terdapat perbedaan yang signifikan pada hasil koefisien siluet antara penggunaan teknik outliers dan tidak menggunakan teknik outliers.

Tabel 12 Hasil Sebelum memakai Outliers
<b>Silhouette Coefficient For The Above Clustering</b>
0.6822638489278713

Tabel 13 Hasil Sesudah memakai Outliers
<b>Silhouette Coefficient For The Above Clustering</b>
0.6454304191549617

Tabel 12 menunjukkan hasil nilai koefisien siluet dengan adanya outliers, sedangkan pada Tabel 13 menunjukkan hasil koefisien siluet setelah menggunakan teknik penghilangan outliers. Terdapat sedikit perubahan dalam hasil koefisien siluet setelah penghilangan outliers, karena adanya outlier yang memengaruhi hasil sebelumnya. Oleh karena itu, nilai yang memanfaatkan metode outliers cenderung lebih rendah dari pada yang tidak menggunakan outliers, karena hal ini mempengaruhi cara outliers tersebut divisualisasikan. Namun, perubahan nilai silhouette coefficient bukan berarti kualitas pengelompokan menurun.

### H. Hasil

Dalam perbandingan antara penggunaan teknik penghilang outliers dan tanpa teknik penghilang outliers, penggunaan teknik penghilang outliers lebih efektif dalam menghasilkan kelompok yang signifikan. Untuk pengelompokan akhir terdiri dari tiga kelompok, yaitu kelompok rekanan potensial tinggi dengan 178 rekanan, kelompok rekanan potensial sedang dengan 473 rekanan, dan kelompok rekanan potensial rendah yang terdiri dari 119 rekanan vendor. Hasil evaluasi menunjukkan nilai 0.6454 dari koefisien silhouette, yang mengindikasikan kualitas kluster yang baik.

## IV. KESIMPULAN

Berdasarkan hasil dan analisis yang telah dilakukan pada penelitian ini, dapat disimpulkan bahwa perbandingan dua teknik dalam mengelompokkan data, yaitu dengan membiarkan outliers dalam data atau menghapusnya, menunjukkan bahwa adanya outliers pada kumpulan data akan berdampak negatif pada kualitas cluster yang dihasilkan. Oleh karena itu, penting untuk mencari dan menghapus outliers sebelum melakukan pengelompokan data yang besar, sehingga dapat membentuk cluster yang berkualitas baik dan meningkatkan efisiensi dalam menjelajahi kumpulan data. Hasil pengelompokan pada penelitian ini menghasilkan tiga kelompok, yaitu 178 rekanan potensial tinggi, 473 rekanan potensial sedang, dan 119 rekanan potensial rendah, menggunakan algoritma k-medoids dengan nilai silhouette coefficient sebesar 0.6454.

## PENGAKUAN

Naskah Ilmiah ini adalah sebagian dari penelitian Tugas Akhir milik Nopiti Yulistiani, dengan judul “ Pengaruh Outlier pada Algoritma K-Medoid untuk Mengelompokkan Rekanan Vendor dalam Pengadaan Barang” yang dibimbing oleh Ayu Ratna Juwita dan Anis Fitri Nur Masruriyah.

## DAFTAR PUSTAKA

- [1] Agustina, N., & Prihandoko, P. (2018). Perbandingan Algoritma K-Means dengan Fuzzy C-Means Untuk Clustering Tingkat Kedisiplinan Kinerja Karyawan. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 2(3), 621–626. <https://doi.org/10.29207/resti.v2i3.492>
- [2] Apriyaningsih, A. (2022). Sistem Pendukung Keputusan Pemilihan Vendor Project Menerapkan Metode Analytical Hierarchy Process (AHP). *TIN: Terapan Informatika Nusantara*, 2(9), 542–550. <https://doi.org/10.47065/tin.v2i9.1309>
- [3] Arif, M. (n.d.). Penerapan Aplikasi Machine Learning Untuk Optimasi Key Performance Indicator ( KPI ) Pada Layanan Jaringan LTE.
- [4] Aulia, S. (2021). Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja). *Djtechno: Jurnal Teknologi Informasi*, 1(1), 1–5. <https://doi.org/10.46576/djtechno.v1i1.964>.
- [5] Aziz, A., Siregar, A., & Zonyfar, C. (2022). Penerapan Algoritma K-Means dan Fuzzy C-Means untuk Pengelompokan Kabupaten Kota Berdasarkan Produksi Padi di Provinsi Jawa Barat. ... *Student Journal for ...*, III, 1–8. <https://journal.ubpkarawang.ac.id/mahasiswa/index.php/ssj/article/view/411>.
- [6] Azizah, A. N., Santoso, K. A., & Jember, U. (2022). PENENTUAN LOKASI ATM BANK SYARIAH INDONESIA MENGGUNAKAN METODE FUZZY C MEANS DI KABUPATEN JEMBER. *Jurnal Matematika, Sains, dan Teknologi*, 23, 12–23.