

Clustering of Accounts Receivable Billing Data Based on Customer Tariff Categories at PT PLN UP3 Palembang

Dimaz Gymnastiar Ramadhan¹, Yulistia²

^{1,2} Universitas Multi Data Palembang, Kota Palembang – Sumatera Selatan, 30113, Indonesia

Informasi Artikel

Diterima : 21 Februari 2025
Revisi : 25 April 2025
Publikasi : 20 Juni 2025

Kata Kunci:

Klaster,
Knowledge Discovery in Database
(KDD),
K-means,
Silhouette Score

ABSTRAK

Penelitian ini bertujuan untuk mengelompokkan pelanggan berdasarkan pola keterlambatan pembayaran dengan menerapkan algoritma *K-Means Clustering*. Data yang digunakan yaitu data piutang terlambat dan tunggakan pelanggan PT PLN Palembang. Adapun metodologi yang dilakukan dalam penerapan data mining ini yaitu tahap pengumpulan data, *preprocessing* menggunakan metode *Knowledge Discovery in Database* (KDD), pemodelan hingga pengujian pengujian aplikasi. Hasil penelitian ini menunjukkan bahwa Klaster 1 memiliki 10 data, Klaster 2 memiliki 36 data, dan Klaster 3 memiliki 326 data terhadap keterlambatan pembayaran. Sedangkan pada tunggakan pembayaran yang berisiko terdapat Klaster 1 memiliki 26 data, Klaster 2 memiliki 36 data, dan Klaster 3 memiliki 312 data. Dari hasil evaluasi menggunakan *Silhouette Score* menunjukkan bahwa 3 Klaster dengan nilai 0,880 (Tertinggi), yang berarti *Clustering* yang dibentuk telah berhasil dan dapat digunakan.

ABSTRACT

The purpose of writing this final assignment is to group customers based on late payment patterns by applying the K-Means Clustering algorithm. The data used are late receivables and arrears of PT PLN Palembang customers. The methodology used in implementing this data mining is the data collection stage, preprocessing using the Knowledge Discovery in Database (KDD) method, modeling to testing the application. The results of writing this final assignment show that Cluster 1 has 10 data, Cluster 2 has 36 data, and Cluster 3 has 326 data on late payments. While in the risky payment arrears, Cluster 1 has 26 data, Cluster 2 has 36 data, and Cluster 3 has 312 data. From the evaluation results using Silhouette Score, it shows that there are 3 clusters with a value of 0,880 (Highest), which means that the clustering that was formed was successful and can be used.

This is an open-access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license



*Penulis Koresponden

Email: dimazgymnr@mhs.mdp.ac.id

Cara sitasi IEEE::

D.G. Ramadhan, T. Prayono, dan Yulistia, "Clustering of Accounts Receivable Billing Data Based on Customer Tariff Categories at PT PLN UP3 Palembang" *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol. 5, no. 2, pp. 394-402, Juni 2025. doi: 10.30811/jaise.v5i2.6511

1. PENDAHULUAN

Di era digital saat ini, pengolahan data pembayaran listrik menjadi elemen krusial dalam mendukung efisiensi dan efektivitas layanan perusahaan penyedia listrik. Di Indonesia, PT. PLN menjadi satu-satunya perusahaan yang bertanggung jawab dalam penyediaan dan pengelolaan distribusi listrik bagi masyarakat, khususnya di Kota Palembang. Selain memberikan layanan kepada pelanggan, PT. PLN Palembang juga bertugas untuk mengawasi kepatuhan pembayaran tagihan listrik dan menerapkan tindakan terhadap pelanggan yang terlambat hingga menunggak [1].

Meskipun PT. PLN Palembang telah melakukan berbagai upaya dalam mengelola tagihan piutang pelanggan, perusahaan masih menghadapi tantangan signifikan terkait keterlambatan pembayaran listrik. Untuk menangani hal ini, PT. PLN membentuk tim penagihan yang bertugas untuk memberikan tindakan kepada pelanggan yang melewati jatuh tempo, mulai dari pengiriman invoice pertama hingga tindakan pencabutan kabel listrik bagi yang tidak melunasi tagihan dalam jangka waktu tiga bulan. Pendekatan ini menyebabkan biaya operasional yang cukup besar, terutama terkait tenaga kerja dan kebutuhan logistik. Oleh karena itu, PT. PLN perlu mengevaluasi data tagihan piutang serta pola keterlambatan pembayaran guna merumuskan strategi yang lebih efisien dan meningkatkan kepatuhan pelanggan. Selain itu, diperlukan sosialisasi yang intensif untuk mendorong disiplin pelanggan dalam membayar tepat waktu, demi membantu mengurangi biaya operasional, hingga risiko kerugian bagi perusahaan.

Salah satu pendekatan yang efektif dalam mengatasi permasalahan keterlambatan dan tunggakan pembayaran pelanggan adalah dengan menerapkan teknik *data mining*. Melalui data mining, PT. PLN dapat melakukan segmentasi pelanggan secara lebih jelas dan sistematis, sehingga mampu merancang strategi penagihan yang lebih terarah dan efektif. Implementasi data mining tidak hanya berfungsi untuk mengelompokkan pelanggan berdasarkan pola tertentu dalam data historis, tetapi juga memberikan wawasan yang berharga untuk memprediksi perilaku pelanggan, meningkatkan efisiensi operasional, dan mengoptimalkan alokasi sumber daya perusahaan [2].

Beberapa penelitian sebelumnya telah membuktikan efektivitas penerapan data mining menggunakan algoritma *K-Means Clustering* dalam berbagai konteks. Salah satunya adalah penelitian berjudul "*Penerapan Data Mining dalam Pengelompokan Buku yang Dipinjam Menggunakan Algoritma K-Means*" [2], yang berhasil mengidentifikasi kategori buku yang paling diminati berdasarkan jumlah ketersediaan dan frekuensi peminjaman, sehingga dapat membantu penyedia layanan perpustakaan dalam pengelolaan koleksi buku secara lebih akurat. Penelitian lain berjudul "*Pengelompokan Data Penerimaan Pajak Bumi dan Bangunan Berdasarkan Kelurahan di Kota Tasikmalaya Menggunakan Algoritma K-Means*" [3] memanfaatkan metode ini untuk mengetahui kelompok penerimaan pajak terendah dan tertinggi guna menentukan prioritas penagihan.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan algoritma *K-Means Clustering* dalam proses pengelompokan pelanggan di PT. PLN Palembang berdasarkan kesamaan karakteristik pembayaran mereka. Dengan demikian, perusahaan dapat mengidentifikasi kelompok pelanggan berisiko tinggi terhadap keterlambatan dan tunggakan pembayaran, sekaligus merancang strategi penagihan yang lebih tepat sasaran. Selain itu, hasil pengelompokan ini diharapkan dapat membantu perusahaan dalam meminimalisasi biaya operasional penagihan dan meningkatkan efektivitas proses penagihan piutang pelanggan.

2. METODE

2.1 Data Mining

Data mining adalah proses mengungkap pengetahuan atau informasi berharga dari kumpulan data yang sangat besar dan kompleks. Tujuan utama dari *data mining* adalah untuk menemukan pola, hubungan, atau informasi yang tidak terlihat secara langsung dalam data tersebut, sehingga bisa memberikan wawasan yang lebih mendalam dan bernilai [4]. Hasil dari proses ini dapat dimanfaatkan sebagai dasar dalam pengambilan keputusan yang lebih cepat, tepat, dan strategis, khususnya dalam dunia bisnis [5].

2.2 Clustering

Clustering adalah metode data mining tanpa supervisi yang bertujuan menemukan pola atau struktur dalam data yang tidak berlabel. Proses ini membantu mengelompokkan data ke dalam grup yang berbeda serta mengurangi dimensi data. Terdapat berbagai jenis metode *clustering*, seperti *hierarchical*, *partitional*, *grid*, *density-based*, dan *model-based*, dengan performa yang bervariasi tergantung pada jenis dan jumlah data yang digunakan [6]. Hasil dari algoritma ini juga dapat membantu dalam mengurangi dimensi data.

Tujuan utamanya adalah untuk mengidentifikasi kelompok-kelompok yang berbeda di dalam suatu himpunan data [7].

2.3 Algoritma K-Means

K-Means adalah metode pengelompokan data (*clustering*) yang bekerja dengan membagi data ke dalam sejumlah kelompok berdasarkan kedekatan data terhadap titik tengah (*centroid*) masing-masing kelompok, dan proses ini dilakukan secara berulang hingga terbentuk kelompok yang stabil [8]. Prinsip dasar algoritma ini adalah meminimalkan jumlah kuadrat kesalahan (*Sum of Squared Error/SSE*) antara objek data dan *centroid* yang ditentukan. Proses kerja algoritma ini terdiri dari empat langkah. Pertama, sejumlah k *centroid* dipilih secara acak dari data yang akan dikelompokkan. Kemudian, setiap objek data yang bukan *centroid* dimasukkan ke dalam kluster yang terdekat dengan *centroid* tersebut. Proses ini diulang (iterasi) hingga posisi *centroid* tidak berubah lagi, yaitu ketika *centroid* yang dihasilkan pada iterasi saat ini sama dengan *centroid* pada iterasi sebelumnya, yang berarti algoritma telah mencapai titik stabil atau konvergen [9].

2.4 Teknik Pengumpulan Data

2.4.1 Data Primer

1) Studi Pustaka

Studi pustaka dilakukan dengan menelusuri, mengutip, dan mencatat referensi dari berbagai buku, hasil penelitian sejenis dengan penelitian mengenai implementasi *K-Means* dalam sistem pengelompokan.

2) Wawancara

Wawancara ini bertujuan untuk memperoleh pemahaman terkait prosedur pengelolaan data piutang pelanggan dan kebijakan perusahaan dalam penanganan tagihan.

2.4.2 Data Sekunder

Tahap awal dalam penelitian ini adalah pengumpulan data, yang dilakukan melalui kerjasama dengan petugas di PT. PLN Palembang. Data-data ini akan digunakan sebagai atribut dalam menganalisis dan memprediksi tingkat risiko keterlambatan hingga tunggakan pembayaran listrik di PT. PLN.

2.5 Perancangan Metode

Data yang digunakan dalam penelitian ini adalah satu dataset yang berisi informasi tagihan piutang pelanggan berdasarkan golongan tarif dengan lama periode 10 bulan, dengan total 44 data golongan tarif dan 396 record transaksi. Adapun deskripsi variabel dataset yang telah diambil dapat dilihat sebagai berikut.

Tabel 1. Deskripsi Variabel Data

No	Atribut	Transaksi	Tipe Data	Deskripsi
1.	Periode (Bulan)	Transaksi	<i>Object</i>	Januari 2024 s.d Oktober 2024
2.	Jenis Pelanggan		<i>Object</i>	Sosial, Rumah Tangga, Bisnis, Industri, Kantor Pemerintah dan Layanan Khusus yang berskala kecil hingga besar
3.	Kode Golongan		<i>Object</i>	S1, S2, S3, R1, R2, R3, B1, B2, B3, I1, I2, I3, I4, P1, P2, T dan C/TM
4.	Tarif Daya		<i>Object</i>	220VA, 450VA, 900VA, 1300VA, hingga 2200VA
5.	Total Pelanggan		<i>Int</i>	Total keseluruhan pelanggan per golongan tarif daya
6.	Jumlah Pelanggan (Tepat Waktu)	Pelanggan	<i>Int</i>	Sejumlah pelanggan yang tepat waktu membayar tagihan
7.	Tagihan Rupiah (Tepat Waktu)		<i>Float</i>	Total biaya tagihan pelanggan yang tepat waktu
8.	Jumlah Pelanggan (Terlambat)	Pelanggan	<i>Int</i>	Sejumlah pelanggan yang terlambat membayar tagihan
9.	Tagihan Rupiah (Terlambat)	Rupiah	<i>Float</i>	Total biaya tagihan pelanggan yang terlambat
10.	Jumlah Pelanggan	Pelanggan	<i>Int</i>	Sejumlah pelanggan yang

11.	(Menunggak) Tagihan Rupiah (Menunggak)	Float	menunggak membayar tagihan Total biaya tagihan pelanggan yang menunggak
-----	--	-------	---

Pada tahap ini penulis menggunakan Proses *Knowledge Discovery in Databases* (KDD). Proses tersebut terdiri dari:

1) *Data Cleaning*

Proses data mining dapat dilakukan setelah melakukan proses pembersihan pada pemilihan data yang menjadi fokus dalam *Knowledge Discovery in Databases* (KDD). Proses pembersihan ini meliputi penghapusan data duplikat, pemeriksaan terhadap data yang tidak konsisten, serta perbaikan kesalahan dalam data seperti kesalahan pengetikan (tipografi) [10]. Untuk itu, penulis menggunakan *Google Colab* sebagai tools yang membantu proses pembersihan data.

2) *Data Selection*

Tahap *data selection* sangat penting sebelum proses transformasi, karena seleksi atribut yang tepat akan menentukan informasi yang dapat diperoleh dari proses tersebut. Pada penelitian ini, dipilih 6 dari 10 atribut yang dinilai paling relevan terhadap tujuan segmentasi pelanggan berdasarkan perilaku pembayaran. Atribut Kode Gol dan Tarif Daya digunakan untuk mengidentifikasi jenis pelanggan dan kapasitas daya yang memengaruhi pola pembayaran. Sementara itu, Persentase Terlambat, Tagihan Terlambat (Rp), Persentase Menunggak, dan Tagihan Menunggak (Rp) dipilih karena langsung merepresentasikan frekuensi serta nominal tagihan pembayaran.

3) *Transformation*

Selanjutnya penulis melakukan transformasi data, dengan menambahkan kolom baru seperti persentase pelanggan terlambat dan menunggak. Penambahan kolom ini dilakukan untuk memberikan gambaran yang lebih proporsional mengenai tingkat keterlambatan pembayaran di setiap golongan tarif daya. Persentase pelanggan terlambat atau menunggak tersebut dapat dihitung menggunakan rumus :

$$\text{Persentase Pelanggan Terlambat} = \frac{\text{Jumlah Pelanggan Terlambat}}{\text{Jumlah Pelanggan Asli}} \times 100$$

Selain itu, untuk memastikan skala data yang seragam dalam proses clustering, dilakukan normalisasi data menggunakan teknik *Min-Max Scaling*. Normalisasi ini bertujuan untuk mengubah nilai data menjadi rentang 0 hingga 1, sehingga setiap variabel memiliki bobot yang setara [11]. Rumus normalisasi *Min-Max* adalah:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Setelah normalisasi data ke variabel yang akan digunakan telah selesai, maka data siap dilanjutkan ke tahap data mining agar penulis dapat melakukan perhitungan algoritma *K-Means Clustering*. Berikut data yang telah diubah menggunakan data yang telah dibersihkan

Tabel 1. Transformasi dan Normalisasi Data

No	Kd_Gol	Tarif Daya	Persentase Terlambat	Tagihan Terlambat (Rp)	Persentase Menunggak	Tagihan Menunggak (Rp)
1	S.2	450 VA	0,448836119	0,002776118	0,020056627	0,000321822
2	S.2	900 VA	0,500223714	0,010333199	0,018139522	0,000851832
3	S.2	1.300 VA	0,363329459	0,008349136	0,022437231	0,002041148
4	S.2	2.200 VA	0,276259866	0,005182065	0,02030784	0,001539659
5	S.2	3.500 VA s/d 200 KVA	0,187476889	0,076785137	0,031023938	0,02525917
6	S.3	200	0,83615819	0,023705442	0,028631238	0,015234992

		KVA				
7	R.1	450 VA	0,479855368	0,371837824	0,019281445	0,042859823
...
337	R.1	450 VA	0,430018832	0,333756341	0,048875013	0,137935655

4) Data Mining

Setelah data melewati proses *pre-processing* hingga transformasi dengan baik, maka tentunya data tersebut siap untuk diolah dalam tahap *data mining* yaitu pemodelan. Salah satu algoritma yang digunakan penulis adalah algoritma K-Means. Menurut [12], Algoritma *K-means Clustering* adalah salah satu metode analisis kluster non-hirarki yang berfungsi mempartisi data ke dalam bentuk satu atau lebih cluster atau kelompok.

Langkah-langkah untuk melakukan *K-Means* adalah sebagai berikut:

- Menentukan jumlah *K* (*cluster*).
- Inisialisasikan *Cluster Centroids* (rata-rata) atau “*mean*” secara acak *Centroid* adalah titik pusat setiap cluster.
- Menghitung jumlah jarak antara data dan centroid dengan menggunakan persamaan *Euclidean Distance*:

$$d(P, Q) = P_1 - Q_{10} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

- Kelompokkan data ke dalam cluster dengan jarak yang paling dekat atau minimum dari setiap data dengan centroid.
- Menghitung nilai centroid yang baru menggunakan rumus :

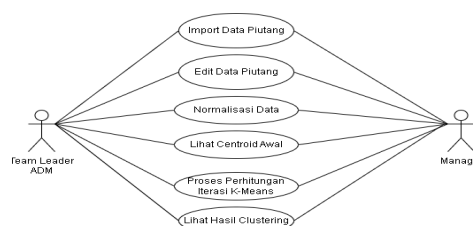
$$\text{Centroid Baru} = \frac{\text{menjumlahkan semua nilai pada setiap cluster yang sama}}{\text{Jumlah data yang ada pada cluster tersebut}}$$

Ulangi langkah ketiga sampai kelima sehingga sudah tidak ada lagi data yang berpindah ke cluster lain

3. HASIL DAN PEMBAHASAN

3.1. Analisis Use Case Diagram

Setelah melakukan pengolahan terhadap dataset yang ada, langkah selanjutnya adalah melakukan analisis kebutuhan sistem menggunakan *use case diagram*, sehingga kebutuhan fungsional dari sistem yang akan diimplementasikan dapat lebih jelas digambarkan. Adapun *use case diagram* yang dapat menggambarkan proses tersebut dapat dilihat pada gambar dibawah ini :



Gambar 3. 1 Use Case Diagram

3.2 Pemodelan

3.2.1. Perhitungan Manual K-Means

Iterasi Ke-1 :

- Menentukan pusat awal *Cluster*

Sebelum menentukan pusat awal *cluster*, terlebih dahulu dilakukan penentuan jumlah cluster (*k*). Pada penelitian ini, jumlah cluster ditentukan secara trial-error, yaitu dengan mencoba beberapa nilai *k* dan menganalisis hasil klusterisasi yang dihasilkan. Setelah beberapa

kali pengujian, jumlah *cluster* yang paling sesuai dengan karakteristik data dan kebutuhan analisis adalah sebanyak 3 *cluster*. Selanjutnya, nilai random awal centroid ditentukan secara acak sesuai jumlah *cluster* yang telah dipilih, sehingga dapat dilihat pada tabel dibawah ini:

Tabel 3. Titik Pusat *Cluster* Awal

C	Kode Gol.	Tarif Daya	Persentase Terlambat	Tagihan Terlambat(Rp)
C1	P.1	6.600 VA s/d 200 kVA	0,700124378	0,484468283
C2	R.1	900 VA	0,451510166	0,221132176
C3	I.3	> 200 kVA	0,011373578	0,001595361

b. Perhitungan jarak pusat *cluster*

Untuk mengukur jarak antara data dengan pusat *cluster*, maka digunakan Euclidean distance dengan rumus sebagai berikut :

Berikut contoh perhitungan jarak dari data ke-1 terhadap pusat *cluster*:

$$d(x_1, c_1) = \sqrt{(0,700124378 - 0,44836119)^2 + (0,484468283 - 0,002776118)^2} = 0,543298$$

$$d(x_1, c_2) = \sqrt{(0,451510166 - 0,44836119)^2 + (0,221132176 - 0,002776118)^2} = 0,218372$$

$$d(x_1, c_3) = \sqrt{(0,011373578 - 0,44836119)^2 + (0,001595361 - 0,002776118)^2} = 0,437464$$

Dan seterusnya dilanjutkan untuk data ke 2, 3,...n. Kemudian akan didapatkan matrik jarak untuk iterasi ke-1.

c. Pengelompokan Data

Jarak dari hasil perhitungan yang telah dilakukan, akan dibandingkan dan dipilih jarak terdekat atau nilai terkecil dari ketiga hasil perhitungan jarak setiap *cluster*, jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan *cluster* terdekat [13].

Tabel 4. Pengelompokan Data

No	C1	C2	C3
1.		*	
2.		*	
3.		*	
4.			*
5.			*
6.			*
7.		*	
8.		*	
9.	*		
10.			*
..
337.		*	

d. Perhitungan Titik *Centroid* Baru

Setelah mendapatkan *cluster* pada iterasi pertama, penulis menentukan kembali nilai *centroid* baru berdasarkan data golongan tarif tiap *cluster* yang sudah didapatkan, sehingga hasil perhitungan nilai *centroid* awal pada persentase terlambat dan tagihan terlambat selanjutnya adalah sebagai berikut :

Tabel 1. Pusat *Cluster* Baru

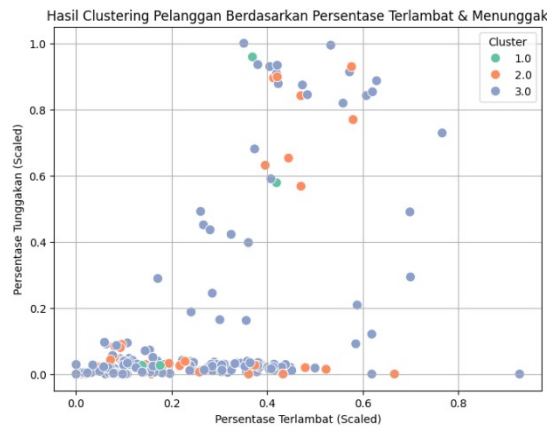
C	Persentase Terlambat	Tagihan Terlambat(Rp)
C1	0,407190505	0,727054208
C2	0,421	0,0663
C3	0,073130193	0,024034326

Ulangi langkah kedua hingga data tidak mengalami perubahan. Langkah selanjutnya yaitu melakukan perbandingan-perbandingan hasil *cluster* pada iterasi pertama dengan kedua. Jika nilai anggota *cluster* memiliki nilai yang konvergen/sama maka proses iterasi berhenti. Iterasi dihentikan dan hasil *cluster* tersebut merupakan hasil *cluster* yang sudah *valid*.

Tabel 6. Perbandingan Iterasi

No	Iterasi 1	Iterasi 2
1.	C2	C2
2.	C2	C2
3.	C2	C2
4.	C3	C2
5.	C3	C3
6.	C3	C3
7.	C2	C2
8.	C2	C2
9.	C1	C1
10.	C3	C3
..
337.	C2	C2

Proses clustering ini memerlukan total 3 iterasi untuk kategori terlambat dan 4 iterasi untuk kategori menunggak guna mencapai hasil yang stabil dan valid. Pada akhir proses, terbentuklah 3 cluster yang merepresentasikan kategori risiko keterlambatan dan tunggakan pembayaran piutang berdasarkan golongan tarif pelanggan. Dari total 374 data piutang yang dianalisis, didapatkan hasil karakteristik tiap cluster sebagai berikut:



Gambar 3.2 Visualisasi Hasil *Clustering*

Tabel 7. Distribusi Golongan Tarif

Cluster	Gol Tarif Berisiko Terlambat	Gol Tarif Berisiko Menunggak	Keterangan
1.	R.1M - 900 VA	P.1 - 6.600 VA s/d 200 kVA	Perlu perhatian khusus
2.	R.1 - 1.300 VA	P.1 - 1.300 VA	Monitor lebih lanjut
3.	B.1 - 1.300 VA	R.1 - 900 VA	Perlu pengawasan ringan

Berdasarkan hasil *clustering* dan tabel distribusi golongan tarif pada masing-masing cluster, terlihat bahwa setiap cluster didominasi oleh golongan tarif tertentu. Hal ini menunjukkan bahwa karakteristik pelanggan berdasarkan daya listrik memiliki pengaruh signifikan terhadap risiko keterlambatan dan tunggakan pembayaran, sehingga segmentasi pelanggan berdasarkan golongan tarif dapat menjadi dasar yang kuat dalam menyusun strategi penagihan yang lebih terarah dan efektif.

3.3 Evaluasi Model

Dalam mengevaluasi kinerja algoritma *K-Means* yang digunakan, penulis mendapatkan hasil perhitungan *Silhouette Score* yang diperoleh menggunakan perintah di *Google Colab*. *Silhouette Score* adalah metrik untuk menilai kualitas clustering dengan membandingkan jarak antar data dalam satu cluster dan jarak ke *cluster* terdekat. Nilai yang lebih tinggi menunjukkan hasil pengelompokan yang lebih baik [14]. Metode ini mampu memberikan informasi terkait kualitas hasil klusterisasi [15]. Berdasarkan hasil pengujian *Silhouette Score*, dapat disimpulkan bahwa pemilihan jumlah kluster sebanyak 3 merupakan pilihan yang optimal dalam membentuk kelompok-kelompok yang homogen pada dataset. Nilai *Silhouette Score* tertinggi sebesar 0,880 menunjukkan bahwa pengelompokan data menjadi tiga kelompok tersebut lebih baik.

4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, penulis berhasil mengimplementasikan metode *data mining* untuk mengelompokkan data piutang terhadap golongan tarif pelanggan berdasarkan tingkat keterlambatan dan tunggakan pembayaran. Data yang digunakan meliputi periode Januari hingga Oktober 2024 dengan total 374 record. Melalui pemodelan *clustering*, data piutang pelanggan berhasil dikelompokkan menjadi tiga kelompok, yaitu kelompok dengan risiko tinggi (C1), sedang (C2), dan rendah (C3).

Hasil pengelompokan menunjukkan bahwa pada data keterlambatan, kelompok C1 didominasi golongan R.1M - 900 VA dari 10 data, C2 oleh golongan R.1 - 1.300 VA dari 38 data, dan C3 oleh golongan B.1 - 1.300 VA dari 326 data. Sedangkan pada data tunggakan, kelompok C1 didominasi golongan P.1 - 6.600 VA s/d 200 kVA dari 26 data, C2 oleh golongan P.1 - 1.300 VA dari 36 data, dan C3 oleh golongan R.1 - 900 VA dari 312 data. Selanjutnya, pengujian dengan menggunakan *Silhouette Score* menghasilkan nilai 0,88812, yang menunjukkan bahwa hasil *clustering* yang dibentuk sudah sangat baik dan dapat digunakan sebagai dasar untuk pengambilan keputusan dalam aplikasi manajemen piutang pelanggan.

Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi yang signifikan dalam meningkatkan efisiensi penanganan piutang pelanggan melalui penerapan data mining, yang dapat membantu perusahaan dalam mengidentifikasi kelompok pelanggan dengan tingkat risiko yang berbeda dan merumuskan strategi penanganan yang lebih efektif.

REFERENSI

- [1] M. N. Ma'rif, "Pengaruh Pelaksanaan Kebijakan Listrik Subsidi Terhadap Manajemen Pengaduan Keluhan Dalam Meningkatkan Efektivitas Pelaksanaan Program Pengaduan Keluhan Di Pt. PLN (Persero) Area Garut," *J. Publik*, vol. 15, no. 1, hal. 35–48, 2021, doi: 10.52434/jp.v15i1.47.
- [2] H. L. Sari dan I. Y. Beti, "Penerapan Data Mining Dalam Pengelompokan Buku Yang Dipinjam Menggunakan Algoritma K-Means," *Kaji. Ilm. Inform. dan Komput.*, vol. 3, no. 6, hal. 925–933, 2023, doi: 10.30865/klik.v3i6.826.
- [3] R. Nursaniah, N. Rahaningsih, I. Ali, dan N. Dienwati Nuris, "Pengelompokan Data Penerimaan Pajak Bumi Dan Bangunan Berdasarkan Kelurahan Di Kota Tasikmalaya Menggunakan Algoritma K-Means," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 2, hal. 1477–1483, 2024, doi: 10.36040/jati.v8i2.9009.
- [4] P. W. Rahayu dkk., *Buku Ajar Data Mining*. Jambi: PT. Sonpedia Publishing Indonesia, 2024.
- [5] Normah, B. Rifai, S. Vambudi, dan R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *J. Tek. Komput. AMIK BSI*, vol. 7, no. 2, hal. 174–180, 2021, doi: 10.31294/jtk.v4i2.
- [6] M. Faizan, M. F. Zuhairi, S. Ismail, dan S. Sultan, "Applications of Clustering Techniques in Data Mining: A Comparative Study," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 12, hal. 146–153, 2020, doi: 10.14569/IJACSA.2020.0111218.
- [7] J. L. M. Rocha, M. A. C. Zela, N. I. V. Torres, dan G. S. Medina, "Analogy of the Application of Clustering and K-Means Techniques for the Approximation of Values of Human Development Indicators," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 9, hal. 526–532, 2021, doi: 10.14569/IJACSA.2021.0120959.
- [8] S. S. Nagari dan L. Inayati, "Implementation of Clustering Using K-Means Method To Determine Nutritional Status," *J. Biometrika dan Kependud.*, vol. 9, no. 1, hal. 62–68, 2020, doi: 10.20473/jbk.v9i1.2020.62-68.
- [9] Suyanto, *Data Mining Untuk Klasifikasi dan Klusterisasi Data*. Bandung: Informatika Bandung, 2017.
- [10] N. I. Febianto dan N. Palasara, "Analisa Clustering K-Means Pada Data Informasi Kemiskinan Di Jawa Barat Tahun 2018," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 8, no. 2, hal. 130–140, 2019, doi: 10.32736/sisfokom.v8i2.653.
- [11] Wenny, "Normalisasi Data Kependudukan Dengan Model Min Max Dan Algoritma K-Means Untuk Pengelompokan Tingkat Ekonomi Masyarakat," *Bull. Inf. Syst. Res.*, vol. 2, no. 2, hal. 63–73, 2024, [Daring]. Tersedia pada: <https://journal.grahamitra.id/index.php/bios>
- [12] I. Saputra dan D. A. Kristiyanti, *Machine Learning Untuk Pemula*. Bandung: Informatika Bandung, 2022.
- [13] S. D. K. Wardani, A. S. Ariyanto, M. Umroh, dan D. Rolliawati, "Perbandingan Hasil Metode Clustering K-Means, Db Scanner & Hierarchical Untuk Analisa Segmentasi Pasar," *JIKO (Jurnal Inform. dan Komputer)*, vol. 7, no. 2, hal. 191–201, 2023, doi: 10.26798/jiko.v7i2.796.
- [14] M. A. Ahmed, H. Baharin, dan P. N. E. Nohuddin, "Analysis of K-means, DBSCAN and OPTICS Cluster algorithms on Al-Quran verses," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 8, hal. 248–254, 2020, doi: 10.14569/IJACSA.2020.0110832.
- [15] S. Paembonan dan H. Abduh, "Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat," *PENA Tek. J. Ilm. Ilmu-Ilmu Tek.*, vol. 6, no. 2, hal. 48, 2021, doi: 10.51557/pt_jiit.v6i2.659.