

Segmentasi Profil Risiko Pelanggan Pada Data Accounts Receivable Menggunakan K-Means Clustering Berbasis Elbow Method

¹ Sigit Ananda Murwato, ² Dimas Aditya Putra Wardhana, ³ Siti Clara Baangga Rahawarin, ⁴ Ekstyananda Naufal Fawwaz

^{1,3,4} Program Studi Teknik Informatika, Universitas 17 Agustus 1945 Surabaya

² Program Studi Teknik Robotika dan Kecerdasan Buatan, Universitas 17 Agustus 1945 Surabaya

E-mail: ¹sigitananda@untag-sby.ac.id, ²dimasapw@untag-sby.ac.id,
³rahawarinklara@gmail.com, ⁴ekstyananda@gmail.com

ABSTRAK

Metode konvensional pemantauan piutang usaha menggunakan laporan umur piutang seringkali memiliki keterbatasan karena bersifat statis dan satu dimensi. Pendekatan ini kerap gagal membedakan antara pelanggan dengan volume transaksi besar yang memiliki termin pembayaran khusus, dengan pelanggan yang benar-benar mengalami gagal bayar, sehingga meningkatkan risiko salah sasaran dalam strategi penagihan. Penelitian ini bertujuan untuk mengimplementasikan teknik data mining menggunakan algoritma K-Means Clustering untuk melakukan segmentasi profil risiko pelanggan secara komprehensif dan multidimensi. Penelitian menggunakan data sekunder laporan umur piutang yang melalui tahap feature engineering menjadi tiga dimensi perilaku utama: tingkat keterlambatan (Average Days Past Due), total nilai hutang (Total Debt), dan frekuensi transaksi (Frequency). Algoritma K-Means kemudian diterapkan pada data yang telah dinormalisasi. Untuk menjamin objektivitas pengelompokan, jumlah kluster optimal ditentukan melalui validasi matematis Elbow Method. Evaluasi Elbow Method menunjukkan bahwa variabilitas data paling optimal dibagi menjadi 3 kluster. Hasil segmentasi berhasil mengidentifikasi tiga profil pelanggan, yaitu pelanggan reguler dengan risiko rendah, pelanggan dengan risiko tinggi, dan pelanggan spesial (VIP). Pendekatan klusterisasi dengan data mining terbukti lebih akurat dibandingkan metode tradisional dalam profil risiko pelanggan. Segmentasi ini memberikan wawasan strategis bagi manajemen untuk menerapkan kebijakan penagihan yang tepat sasaran.

Kata kunci: Piutang Usaha, Segmentasi Pelanggan, Data Mining, K-Means Clustering, Elbow Method

ABSTRACT

Traditional Accounts Receivable (AR) monitoring using aging reports is often limited by its static, one-dimensional nature. This approach frequently fails to distinguish between high-volume customers with special payment terms and those with bad debt, thereby increasing the risk of misdirected collection strategies. This study aims to implement data mining techniques using the K-Means Clustering algorithm to perform comprehensive and multidimensional customer risk segmentation. The research uses secondary AR data, which were feature-engineered to derive three key behavioral dimensions: Average Days Past Due (DPD), Total Debt, and Transaction Frequency. The K-Means algorithm was applied to normalized data, and the Elbow Method was used to mathematically determine the optimal number of clusters to ensure objectivity. The

evaluation indicates that the data variability is optimally divided into three clusters. The segmentation results successfully identified three distinct customer profiles: low-risk regular customers, high-risk customers, and special (VIP) customers. This machine learning clustering approach proved more accurate than traditional methods for customer risk profiling. Ultimately, this segmentation provides management with strategic insights to implement targeted and effective collection policies

Keyword: *Accounts Receivable, Customer Segmentation, Data Mining, K-Means Clustering, Elbow Method*

1. PENDAHULUAN

Pengelolaan piutang usaha (*accounts receivable*) memiliki peranan penting dalam menjaga kelancaran likuiditas serta kestabilan arus kas perusahaan. Piutang usaha yang tidak dikelola dengan baik memiliki risiko menjadi piutang yang tidak terbayar seluruhnya atau sebagian. Perputaran piutang yang rendah menyebabkan investasi modal pada piutang menjadi besar dan bertambahnya risiko keterlambatan pembayaran piutang (Dewi Ratna Sari & Sunu Priyawan, 2023). Pendekatan analisis konvensional yang mengandalkan laporan umur piutang (*aging report*) dengan pembagian rentang waktu statis, seperti 0–30 atau 31–60 hari, belum cukup mampu merepresentasikan dinamika dan kompleksitas perilaku pembayaran pelanggan, terutama ketika dihadapkan pada frekuensi transaksi dan skala data yang besar. Kondisi ini menyebabkan kesulitan bagi manajemen dalam melakukan identifikasi risiko pelanggan secara akurat.

Perusahaan harus selalu mampu menjaga umur piutang dalam rentang sependek mungkin. Melakukan segmentasi profil risiko pelanggan berdasarkan umur piutang merupakan langkah awal dalam identifikasi dan mengambil langkah strategi ke depan. Tujuan utama dari segmentasi ini adalah untuk memahami perilaku pelanggan, menganalisis preferensi, dan mengidentifikasi kelompok pelanggan (Ardana et al., 2024). Dengan kondisi jumlah data yang banyak dan kompleks, proses analisis manual menjadi tidak efisien dan rentan terhadap human error. Oleh karena itu, diperlukan pendekatan

komputasi cerdas menggunakan teknik data mining dalam mendapatkan informasi penting untuk membantu proses pengambilan keputusan (Iqbal et al., 2025).

Metode clustering dengan algoritma K-Means dipilih sebagai metode utama karena dalam mengolah data dalam jumlah besar, seringkali kesulitan dalam menentukan target yang menjadi tujuan pengelompokan (Nurhayati et al., 2025). Laporan umur piutang lebih mudah dikategorikan sebagai data yang bersifat *unsupervised*, sehingga metode clustering lebih tepat dan efisien dalam proses pengolahan dan pengelompokan data numerik berskala besar.

Penelitian terdahulu mengenai klasterisasi pelanggan umumnya mengadopsi model RFM (*Recency, Frequency, Monetary*), namun fokus utamanya terbatas pada segmentasi profitabilitas dan retensi pemasaran, bukan pada mitigasi risiko gagal bayar (Christy et al., 2021). Di sisi lain, studi terkait pemodelan risiko kredit umumnya menggunakan pendekatan *supervised learning* untuk persetujuan pinjaman perbankan, yang mana kurang aplikatif untuk dinamika piutang usaha B2B (Koutanaei et al., 2015). Bahkan pada studi yang secara spesifik membahas klasterisasi data piutang, atribut yang digunakan masih bergantung pada nilai faktur tunggal atau umur piutang statis (Gymnastiar Ramadhan & Koresponden, 2025).

Tantangan utama dalam penggunaan metode clustering adalah penentuan jumlah kelompok yang paling optimal (Edy Prayitno et al., 2024). Penentuan

nilai k yang sembarangan dapat menyebabkan hasil segmentasi menjadi bias atau tidak bermakna. Untuk menghasilkan nilai k yang optimal, pada penelitian ini akan menggunakan Elbow Method sebagai validasi matematis untuk menentukan jumlah segmen risiko pelanggan yang paling optimal, sehingga menghasilkan profil risiko yang akurat dan dapat ditindaklanjuti oleh manajemen (Ahsina et al., 2022).

2. LANDASAN TEORI

Piutang Usaha

Piutang usaha (*Accounts Receivable/AR*) didefinisikan sebagai tuntutan pembayaran perusahaan kepada pihak lain (pelanggan) atas penyerahan barang atau jasa yang dilakukan secara kredit. Dalam manajemen keuangan, AR merupakan kelompok aset lancar yang paling likuid setelah kas. Namun, piutang mengandung risiko inherent berupa keterlambatan pembayaran atau gagal bayar (default) (Kurnia et al., 2023).

Secara konvensional, pemantauan risiko piutang dilakukan menggunakan Laporan Umur Piutang (*Aging Schedule*). Metode ini mengklasifikasikan piutang ke dalam interval waktu tertentu berdasarkan tanggal jatuh tempo, umumnya: 0 - 30 hari, 31 - 60 hari, 61 - 90 hari, dan > 90 hari (Fhelly Djun et al., 2024). Kelemahan utama dari metode konvensional ini adalah sifatnya yang agregat dan statis. Analisis umur piutang seringkali gagal mendeteksi anomali perilaku pembayaran pada level individu pelanggan sebelum piutang tersebut menjadi macet (*bad debt*).

K-Means Clustering

K-Means adalah salah satu algoritma clustering berbasis partisi yang paling populer karena kesederhanaan dan efisiensi komputasinya, terutama pada dataset berukuran besar. Prinsip kerja algoritma ini adalah membagi data ke dalam k buah klaster, di mana setiap data akan dimiliki oleh klaster dengan rata-rata

(*mean*) atau pusat (*centroid*) terdekat (Ho et al., 2023; Jordy et al., 2023). Proses ini dilakukan secara iteratif hingga posisi centroid konvergen (tidak berubah lagi).

Evaluasi Clustering

Salah satu tantangan fundamental dalam K-Means adalah menentukan nilai k yang optimal. *Elbow Method* adalah teknik visual untuk mengevaluasi konsistensi internal klaster. Metode ini menghitung nilai WCSS (*Within-Cluster Sum of Squares*), yaitu jumlah kuadrat jarak antara setiap titik data dengan *centroid*-nya. Semakin banyak klaster, nilai WCSS akan semakin kecil (Almeida et al., 2021). Namun, tujuan kita adalah mencari titik "siku" pada grafik, di mana penurunan WCSS mulai melandai secara signifikan. Titik ini dianggap sebagai jumlah klaster yang paling efisien.

Silhouette Coefficient mengukur seberapa mirip suatu objek dengan klasternya sendiri (*cohesion*) dibandingkan dengan klaster lain (*separation*) (Kaur et al., 2023). Nilai *Silhouette (s)* berkisar antara -1 hingga 1:

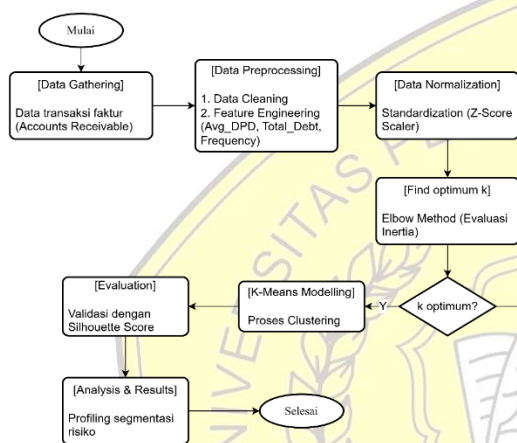
- $s \approx 1$: Data terklusterisasi dengan sangat baik (rapat dan terpisah jauh).
- $s \approx 0$: Data berada di perbatasan antar klaster (*overlapping*).
- $s \approx -1$: Data kemungkinan salah ditempatkan dalam klaster.

3. METODOLOGI

Tahapan Penelitian

Penelitian ini dilakukan melalui beberapa tahapan yang secara umum ilustrasinya ditunjukkan pada Gambar 1. Proses diawali dengan pengumpulan data transaksi piutang usaha yang kemudian dilakukan *data cleaning* dan *feature engineering* menjadi tiga fitur perilaku pelanggan, yaitu *Average Days Past Due (Avg DPD)*, *Total Debt*, dan *Frequency*. Untuk menghindari bias akibat perbedaan rentang nilai yang ekstrem, seluruh fitur tersebut dinormalisasi menggunakan metode *Z-Score Standardization*.

Tahap berikutnya adalah penentuan jumlah kluster optimal menggunakan *Elbow Method* melalui evaluasi nilai Inertia. Setelah jumlah kluster terbaik ditetapkan, algoritma K-Means dijalankan berbasis jarak Euclidean, divalidasi kualitas keterpisahannya dengan Silhouette Score, dan diakhiri dengan interpretasi profil risiko (*Regular, High Risk, dan Special Customers*) sebagai dasar perumusan strategi penagihan perusahaan (Harahap et al., 2021).



Gambar 1. Tahapan Penelitian

Pengumpulan Data

Data yang digunakan adalah laporan umur piutang (*AR Aging Report*) dari perusahaan manufaktur pestisida yang ada di Kabupaten Gresik, Provinsi Jawa Timur. Laporan tersebut berisi data tagihan yang belum dibayar oleh pelanggan dan sudah terklasifikasi berdasarkan rentang waktu periode lama tunggakan pembayaran. Pada laporan umur piutang tersebut memiliki atribut data sebagai berikut:

- CustomerID: Kode identitas pelanggan
- InvoiceNo: Nomor referensi tagihan
- InvoiceDate: Tanggal tagihan
- InvoiceAmount: Nilai nominal tagihan
- Aging (Belum Jatuh Tempo, 0-30, 31-60, 61-90, 90+)

Penyiapan Data (*Data Preprocessing*)

a. Pembersihan Data (*Data Cleaning*)

Pada tahap ini dilakukan proses menyiapkan data agar siap untuk dilakukan analisis. Proses yang dilakukan adalah:

- Penanganan *missing value*: Data transaksi yang tidak memiliki informasi CustomerID atau Invoice_Amount akan dihapus dari dataset.
- Penanganan nilai InvoiceAmount negatif menunjukkan data invoice retur dan akan dihapus dari dataset.

b. Feature Engineering

Beberapa atribut pada laporan umur piutang tersebut akan ditransformasi sesuai dengan kebutuhan dari Algoritma K-Means. Karena data bersifat rincian, yaitu dalam satu pelanggan memiliki banyak tagihan, maka data tersebut akan di-agregat menjadi 1 pelanggan akan terdapat 1 baris data. Kemudian akan dibuat beberapa atribut baru untuk menghasilkan analisa yang lebih tepat:

- Atribut rata-rata berapa hari lama jatuh tempo (*average days past due*): Atribut ini merepresentasikan dimensi risiko dengan melihat bagaimana data tentang keterlambatan bayar.

$$avg_dpd_i = \frac{\sum_{j=1}^n Today_{ij} - InvoiceDate_{ij}}{n_i}$$

Dimana i adalah pelanggan; n adalah jumlah tagihan dari pelanggan tersebut.

- Atribut total nilai tagihan (*total debt*): Atribut ini merepresentasikan dimensi moneter seberapa besar nilai utang dari pelanggan tersebut.

$$total_debt_i = \sum_{j=1}^n InvoiceAmount_{ij}$$

- Atribut total jumlah tagihan yang jatuh tempo (*frequency*): Atribut ini merepresentasikan seberapa aktif pelanggan melakukan transaksi.

$$frequency_i = Count(InvoiceNo)$$

c. Data Normalization

Tiga atribut baru tersebut (*avg_dpd*, *total_debt*, *frequency*) memiliki satuan dan rentang skala yang beragam (Hari, Rupiah dan Frekuensi). Jika kita ambil contoh antara *avg_dpd* dengan *total_debt*, rentang nilainya sangat jauh sehingga hasil analisis dari Algoritma K-Means akan sulit untuk dilakukan. K-Means berbasis jarak (Euclidean Distance) sehingga sangat sensitif terhadap rentang nilai. Maka kemudian diterapkan Z-Score Normalization (Standard Scaler) agar semua fitur memiliki rata-rata = 0 dan standar deviasi = 1.

$$x' = \frac{x - \mu}{\sigma}$$

Metode Clustering (K-Means)

Algoritma K-Means membagi data ke dalam kluster sejumlah k dengan meminimalkan varian di dalam kluster (*Sum of Squared Errors*).

a. Perhitungan Jarak

Seberapa dekat antar data dihitung menggunakan Euclidean Distance dalam ruang n -dimensi:

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

p, q = dua titik dalam ruang Euclidean n dimensi

b. Iterative Algorithm

Proses pada algoritma K-Means dilakukan dengan langkah-langkah sebagai berikut:

- Inisialisasi: Menentukan jumlah k dan memilih k titik *centroid* awal secara acak
- Assignment: Mengalokasikan setiap titik data pelanggan ke kluster dengan *centroid* terdekat.
- Update: Menghitung ulang posisi *centroid* baru berdasarkan rata-rata aritmatika dari semua data dalam kluster tersebut.

- Iterasi: Mengulangi langkah 2 dan 3 hingga posisi *centroid* stabil (konvergen) atau mencapai batas iterasi maksimum.

Penentuan Jumlah Kluster Optimal

Metode penghitungan nilai *Inertia* atau *Within-Cluster Sum of Squares* (WCSS) untuk rentang $k = 1$ hingga $k = 10$:

$$WCSS = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2$$

Nilai k optimal dipilih pada titik siku grafis, yaitu titik di mana penambahan jumlah kluster tidak lagi memberikan penurunan WCSS yang signifikan.

Evaluasi dan Validasi Model

Untuk menjamin validitas hasil pengelompokan secara matematis, digunakan metrik *Silhouette Coefficient*. Metrik ini mengukur seberapa mirip suatu objek dengan klusternya sendiri (*cohesion*) dibandingkan dengan kluster lain (*separation*).

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

Dimana:

- $a(i)$ adalah rata-rata jarak antara data i dengan semua data lain dalam satu kluster.
- $b(i)$ adalah rata-rata jarak antara data i dengan semua data di kluster tetangga terdekat (*nearest neighbour*).
- $-1 > S(i) > 1$; Nilai yang mendekati 1 menunjukkan struktur kluster yang baik.

Lingkungan Pengembangan

Implementasi proses penelitian mulai dari penyiapan data hingga evaluasi model menggunakan bahasa pemrograman Python dan Jupyter Notebook sebagai development environment application. Library machine learning yang digunakan diambil dari scikit-learn, Pandas dan Numpy.

Kemudian untuk library visualisasi menggunakan Matplotlib.

4. HASIL DAN PEMBAHASAN

Data Preprocessing

Keseluruhan record pada data ini berjumlah 1046. Analisis awal pada data agregat menunjukkan sebaran yang tinggi. Nilai maksimum keterlambatan (jatuh tempo) mencapai 2908 hari, dan nilai maksimum utang pelanggan mencapai 27 Miliar rupiah.

	avg_dpd	total_debt	frequency
0	1539	3408350	1
1	490	591963237	18
2	282	2774999	2
3	17	39012187	1
4	243	7779998	2
...	...	...	...
1041	107	3399998	3
1042	215	1599999	1
1043	2818	0	3
1044	315	103563492	7
1045	342	750009	1

[1046 rows x 3 columns]

	avg_dpd	total_debt	frequency
count	1046	1046	1046
mean	452	380501825	4
std	630	1818816317	11
min	3	-203212241	1
25%	141	930625	1
50%	233	3485998	2
75%	319	30599444	3
max	2908	27360170205	208

Gambar 1. Statistik Dataset

Agar atribut utang pelanggan tidak mendominasi pembentukan kluster, maka digunakan *Standard Scaler* untuk menormalisasi data sebelum dilakukan pemodelan.

```
# 2. Scaling (total_debt)
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

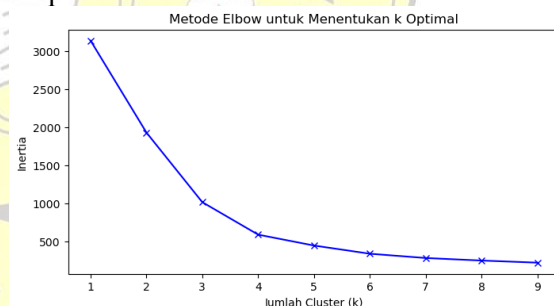
print(X_scaled);
```

```
[[ 1.72508782 -0.20742824 -0.31495793]
 [ 0.05971945  0.11631882  1.26400511]
 [-0.26906913 -0.20777662 -0.22207775]
 ...
 [ 3.75548293 -0.20930307 -0.12919757]
 [-0.21703581 -0.15233578  0.24232314]
 [-0.1746383  -0.20889051 -0.31495793]]
```

Gambar 2. Proses Normalisasi Data

Penentuan K Optimal (Elbow Method)

Berdasarkan pengujian nilai $k = 1$ hingga $k = 10$, grafik *Inertia* menunjukkan penurunan tajam dari $k = 1$ ke $k = 2$, dan mulai melandai secara signifikan pada titik $k = 3$. Penambahan kluster di atas 3 tidak memberikan penurunan variansi yang sepadan dengan kompleksitas model. Sehingga, diputuskan $k = 3$ adalah jumlah kluster optimal.



Gambar 3. Grafik Penentuan k Optimal

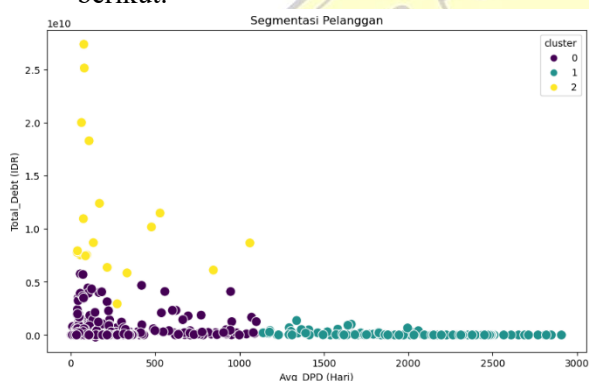
Segmentasi K-Means

Penerapan algoritma K-Means dengan $k = 3$ menghasilkan tiga profil pelanggan dengan analisa karakteristik sebagai berikut.

Tabel 1. Segmentasi Risiko Profil Pelanggan

<i>k</i>	Category	Avg_D PD (hari)	Total Debt (IDR)	Frequency
0	Regular	231 hari	IDR 194.408.887	3
1	High Risk	2023 hari	IDR 84.450.767	4
2	Special Customers	250 hari	IDR 11.185.872.616	60

Dengan menggunakan visualisasi matplotlib, akan didapat hasil sebagai berikut.

**Gambar 4. Hasil Visualisasi K-Means Clustering****Analisa Kategori Klaster**

- Cluster 0: Pelanggan Reguler (Risiko Rendah)
Mayoritas pelanggan berada pada klaster ini, dimana pelanggan memiliki tingkat keterlambatan pembayaran yang rendah dan cenderung tepat waktu. Secara nilai utang pelanggan juga relatif rendah, sehingga pelanggan yang berada pada klaster ini memiliki tingkat risiko yang rendah.
- Cluster 1: Pelanggan Risiko Tinggi
Pelanggan pada klaster ini memiliki tunggakan tagihan yang tak kunjung dibayar. Tagihan yang macet selama bertahun-tahun ini menyebabkan perusahaan tidak dapat menikmati

aset yang likuid (aset lancar) sehingga perusahaan harus segera melakukan tindakan.

- Cluster 2: Pelanggan Berharga
Pelanggan yang berada pada klaster ini merupakan pelanggan yang sering melakukan transaksi dan dengan nominal yang besar. Dan yang lebih penting, tagihan-tagihan pelanggan ini dibayar secara disiplin. Tipe pelanggan seperti ini harus dirawat dengan baik dan diberikan termin khusus agar frekuensi transaksi yang dilakukan lebih sering.

5. KESIMPULAN**Kesimpulan**

Integrasi teknik Data Mining dalam proses pengelolaan piutang usaha memberikan hasil analisis yang presisi yang meningkat secara signifikan dibandingkan dengan metode konvensional. Transformasi data transaksional menjadi atribut multidimensi (*Average Days Past Due*, *Total Debt*, dan *Frequency*) berhasil merepresentasikan kompleksitas perilaku pembayaran pelanggan yang akan sulit ditangkap oleh analisis secara sederhana.

Implementasi algoritma K-Means Clustering pada struktur data tersebut berhasil memetakan pelanggan ke dalam tiga segmen yang berbeda. Hasil klasterisasi ini memberikan landasan strategis yang lebih valid bagi manajemen untuk menerapkan kebijakan penagihan yang tepat, mencegah kesalahan perlakuan terhadap pelanggan prioritas, sekaligus memitigasi risiko *bad debt* lebih dini.

Saran

Perusahaan dapat menggunakan hasil segmentasi ini sebagai fitur standar dalam dashboard monitoring piutang bulanan. Hal ini bertujuan agar manajemen prioritas penagihan tidak lagi bersifat satu dimensi (hanya melihat umur piutang), melainkan bersifat multidimensi dengan mempertimbangkan karakteristik unik

setiap klaster pelanggan, khususnya dalam menangani pelanggan VIP dan berisiko tinggi.

Secara umum, atribut yang digunakan sudah cukup untuk menghasilkan klaster profil risiko pelanggan. Untuk meningkatkan akurasi segmentasi, disarankan untuk mengeksplorasi variabel perilaku tambahan, seperti frekuensi retur atau pola pembayaran parsial. Penelitian selanjutnya dapat menambahkan proses perbandingan performa K-Means dengan algoritma lain yang lebih robust terhadap noise, seperti DBSCAN agar dapat menghasilkan deteksi anomali pelanggan yang lebih spesifik.

6. DAFTAR PUSTAKA

- Ahsina, N., Fatimah, F., & Rachmawati, F. (2022). Analisis Segmentasi Pelanggan Bank Berdasarkan Pengambilan Kredit Dengan Menggunakan Metode K-Means Clustering. *Jurnal Ilmiah Teknologi Infomasi Terapan*, 8(3). <https://doi.org/10.33197/jitter.vol8.iss3.2022.883>
- Almeida, V. C., Rodrigues, B. S., & Silva, H. H. (2021). Data Clusterization Techniques Applied for High Consumption Customer'S Installment Payment Decision. *IET Conference Proceedings*, 2021(6), 3216–3219. <https://doi.org/10.1049/icp.2021.1723>
- Ardana, C. H., Khoyum, A. A. A. A. A., & Faisal, M. (2024). Segmentasi Pelanggan Penjualan Online Menggunakan Metode K-means Clustering. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 9(1), 1–9. <https://doi.org/10.14421/jiska.2024.9.1.1-9>
- Christy, A. J., Umamakeswari, A., Priyatharsini, L., & Neyaa, A. (2021). RFM ranking – An effective approach to customer segmentation. *Journal of King Saud University - Computer and Information Sciences*, 33(10), 1251–1257. <https://doi.org/10.1016/J.JKSUCI.2018.09.004>
- Dewi Ratna Sari, & Sunu Priyawan. (2023). Analisis Efektivitas Manajemen Piutang Terhadap Pengendalian Piutang Tak Tertagih Pada Pt. Sarayu Garuda Elektrindo. *Journal of Student Research*, 1(2), 190–202. <https://doi.org/10.55606/jsr.v1i2.988>
- Edy Prayitno, Perdana, I. J., Iskandar, E., Heri Winarno, B., & Subagyo, A. A. (2024). Optimalisasi Profitabilitas Ritel Melalui Segmentasi Pelanggan Dengan K-Means Clustering. *Informasi Interaktif: Jurnal Informatika Dan Teknologi Informasi*, 9(3). <https://doi.org/10.37159/jii.v9i3.107>
- Fhelly Djun, S., Gede, I., Gunadi, A., & Sariyasa, D. S. (2024). Analisis Segmentasi Pelanggan pada Bisnis dengan Menggunakan Metode K-Means Clustering pada Model Data RFM. *Jurnal Teknologi Informasi Dan Multimedia*, 5(4), 354–364. <https://doi.org/10.35746/JTIM.V5I4.434>
- Gymnastiar Ramadhan, D., & Koresponden, P. (2025). Clustering of Accounts Receivable Billing Data Based on Customer Tariff Categories at PT PLN UP3 Palembang. *Journal of Artificial Intelligence and Software Engineering*, 5(2), 394–402. <https://doi.org/10.30811/JAISE.V5I2.6511>
- Harahap, M., Lubis, Y., & Situmorang, Z. (2021). Data Science bidang Pemasaran : Analisis Prilaku Pelanggan. *Data Sciences Indonesia (DSI)*, 1(1), 21–32. <https://doi.org/10.47709/DSI.V1I1.1194>
- Ho, T., Nguyen, S., Nguyen, H., Nguyen, N., Man, D. S., & Le, T. G. (2023). An Extended RFM Model for Customer Behaviour and Demographic Analysis in Retail Industry. *Business Systems Research*, 14(1), 26–53. <https://doi.org/10.2478/BSRJ-2023-0002>
- Iqbal, I., Hidayat, N., Gevano, D. P., & Ilahi, A. P. R. (2025). Segmentasi Pelanggan Menggunakan K-Means Clustering Berdasarkan Data Kepribadian dan Pola Konsumsi. *Jurnal Teknik Informatika (Jutif)*, 6(5), 3914–3924. <https://doi.org/10.52436/1.JUTIF.2025.6.5.5140>
- Jordy, M., Triayudi, A., & Sholihati, I. D.

- (2023). Analisis Segmentasi Recency dan Customer Value Pada AVANA Indonesia Dengan Algoritma K-Means dan Model RFM (Recency, Frequency and Monetary). *Journal of Information System Research (JOSH)*, 4(2), 579–589.
<https://doi.org/10.47065/JOSH.V4I2.2950>
- Kaur, M., Sharma, A., Kishore, M., Shubhnath, S. S., & Verma, H. (2023). Customer Segmentation with RFM Analysis using ML Models. *3rd International Conference on Innovative Mechanisms for Industry Applications, ICIMIA 2023 - Proceedings*, 743–749.
<https://doi.org/10.1109/ICIMIA60377.2023.10426021>
- Koutanaei, F. N., Sajedi, H., & Khanbabaei, M. (2015). A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring. *Journal of Retailing and Consumer Services*, 27, 11–23.
<https://doi.org/10.1016/J.JRETCONSER.2015.07.003>
- Kurnia, M. D., Tukiyyat, & Miswan. (2023). Klasifikasi Customer Relationship Management Perusahaan Telekomunikasi Seluler Dengan Metode Machine Learning. *Jurnal Penelitian Ilmu Komputer*, 1(4), 63–76.
https://www.researchgate.net/publication/381676282_Klasifikasi_Customer_Relationship_Management_Perusahaan_Telekomunikasi_Seluler_Dengan_Metode_Machine_Learning
- Nurhayati, I. P., Helmayana, Tiani, A., Maruenci, K., Harahap, Y. S., & Fansyuri, M. (2025). Utilization of Data Mining for Credit Card Customer Segmentation Using the K-Means Method. *Journal of Information Technology and Informatics Engineering*, 1(1), 62–66.
<https://journal.jci.co.id/jitie/article/view/162>