

Segmentasi Pelanggan Bisnis Kursus Berbasis LRFM Menggunakan t-SNE, UMAP, dan PaCMAP untuk Optimalisasi Profitabilitas Bisnis

Customer Segmentation for Course-Based Businesses Using LRFM with t-SNE, UMAP, and PaCMAP for Business Profitability Optimization

Yanuar Rafi Rahadian^{*1}, Hasan Bisri², Latifa Indirani Amina³

*Universitas Pembangunan Nasional "Veteran" Jawa Timur, Jalan Raya Rungkut Madya
Surabaya Indonesia, (031)8706369*

*E-mail : yanuar.rafi.ft@upnjatim.ac.id^{*1}, hasan.bisri.ft@upnjatim.ac.id²,
24032010001@student.upnjatim.ac.id*

**Corresponding author*

Received 18 July 2025; Revised 04 August 2025; Accepted 11 August 2025

Abstrak - Segmentasi pelanggan memainkan peran penting dalam meningkatkan profitabilitas bisnis dengan mengidentifikasi pola perilaku pelanggan. Penelitian ini mengusulkan pendekatan segmentasi berbasis model LRFM (Length, Recency, Frequency, Monetary) untuk meningkatkan ketepatan klasifikasi pelanggan. Dengan menggunakan data transaksi, skor LRFM dihitung dan diklasifikasikan ke dalam empat kategori: *Best Customers* (121 pelanggan), *Loyal Customers*, *Potential*, dan *At Risk*. Untuk meningkatkan interpretabilitas hasil, tiga teknik reduksi dimensi : t-SNE, UMAP, dan PaCMAP digunakan dalam analisis visual. Di antara ketiganya, UMAP menunjukkan visualisasi paling seimbang berdasarkan indikator *separation clarity*, *cluster compactness*, dan *outlier identification*. Hasil penelitian menunjukkan bahwa segmentasi visual membantu pengambil keputusan dalam mengidentifikasi kelompok pelanggan strategis untuk program retensi dan optimalisasi. Studi ini memberikan referensi berharga bagi industri yang ingin meningkatkan manajemen hubungan pelanggan melalui pendekatan berbasis data.

Kata kunci - Model LRFM, profitabilitas bisnis, reduksi dimensi, segmentasi pelanggan, UMAP

Abstract — Customer segmentation plays a crucial role in enhancing business profitability by identifying customer behavior patterns. This study proposes an LRFM (Length, Recency, Frequency, Monetary) model-based segmentation approach to improve the accuracy of customer classification. Using transaction data, LRFM scores are calculated and categorized into four groups: *Best Customers* (121 customers), *Loyal Customers*, *Potential*, and *At Risk*. To enhance the interpretability of the results, three dimensionality reduction techniques—t-SNE, UMAP, and PaCMAP—are employed for visual analysis. Among them, UMAP provides the most balanced visualization based on indicators such as *separation clarity*, *cluster compactness*, and *outlier identification*. The findings indicate that visual segmentation assists decision-makers in identifying strategic customer groups for retention programs and optimization efforts. This study offers valuable insights for industries aiming to improve customer relationship management through data-driven approaches.

Keywords — LRFM Model, Business Profitability, Dimensionality Reduction, Customer Segmentation, UMAP

1. PENDAHULUAN

Perusahaan modern menghadapi tantangan yang semakin kompleks dalam mempertahankan pelanggan di tengah persaingan bisnis yang ketat. Salah satu strategi yang krusial dalam meningkatkan profitabilitas adalah memahami perilaku pelanggan melalui

pendekatan segmentasi yang berbasis data. Segmentasi pelanggan memungkinkan perusahaan untuk mengidentifikasi kelompok pelanggan berdasarkan karakteristik tertentu guna menyusun strategi pemasaran yang lebih efektif [1]. Model segmentasi yang banyak digunakan adalah RFM, yaitu *recency*, *frequency*, dan *monetary*, yang mengevaluasi seberapa baru, seberapa sering, dan seberapa besar transaksi yang dilakukan pelanggan. Namun, model RFM memiliki keterbatasan dalam mengukur loyalitas jangka panjang. Untuk mengatasi keterbatasan tersebut, dikembangkan model LRFM dengan menambahkan dimensi *length* yang menggambarkan durasi hubungan pelanggan dengan perusahaan [2]. Segmentasi pasar menjadi semakin kompleks dan krusial karena perusahaan tidak hanya mengandalkan data lokal dan global, namun memanfaatkan informasi pelanggan pada platform digital. Pendekatan ini memungkinkan pemetaan yang lebih relevan untuk pengembangan strategi pemasaran [3]. Model LRFM memberikan pandangan yang lebih utuh terhadap nilai pelanggan, termasuk potensi loyalitasnya. Hal ini menjadikan model LRFM sebagai pendekatan yang lebih unggul dalam klasifikasi pelanggan strategis, seperti pelanggan setia, pelanggan baru, dan pelanggan berisiko [4]. Kombinasi strategi akuisi pelanggan baru dan retensi pelanggan lama sangat penting untuk pertumbuhan bisnis yang berkelanjutan. Manajemen hubungan yang baik akan membantu memahami kebutuhan pelanggan dan memperkuat loyalitas mereka [5]. Meskipun demikian, data hasil pengukuran LRFM bersifat multidimensi dan tidak mudah dianalisis secara visual. Kebutuhan akan visualisasi yang intuitif semakin penting agar pengambil keputusan dapat dengan cepat mengenali pola perilaku pelanggan. Salah satu pendekatan yang digunakan untuk menyelesaikan masalah ini adalah teknik reduksi dimensi, yang menyederhanakan data berdimensi tinggi ke dalam dua atau tiga dimensi.

Metode *t-distributed stochastic neighbor embedding* (t-SNE) merupakan salah satu teknik visualisasi yang mampu mengungkap struktur lokal dalam data pelanggan [6]. Namun demikian, metode *t-distributed stochastic neighbor embedding* (t-SNE) memiliki keterbatasan dalam menjaga struktur global data karena prosesnya bersifat *stochastic* dan sensitif pada parameter hiper [7]. Selain itu, metode *uniform manifold approximation and projection* (UMAP) dikembangkan untuk menjaga keseimbangan antara struktur lokal dan global dalam proyeksi data [8]. Metode terbaru yaitu *pairwise controlled manifold approximation* (PaCMAP) memberikan fleksibilitas dalam mempertahankan hubungan antar data sehingga cocok untuk keperluan visualisasi kluster [9]. Meskipun metode PaCMAP menunjukkan kinerja yang lebih baik, perlu adanya pemilihan metode reduksi dimensi yang berfungsi sebagai eksplorasi data, proses klustersasi seperti K-Means atau DEC dalam menyederhanakan representasi data berdimensi tinggi [10].

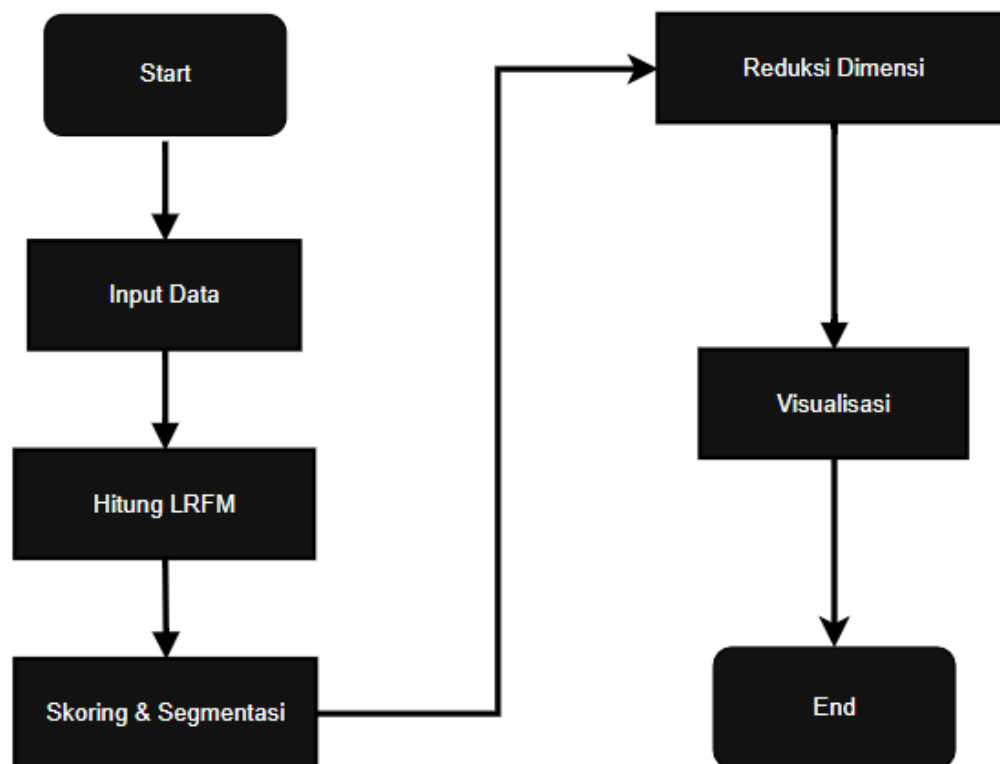
Namun, belum banyak penelitian yang mengintegrasikan model LRFM dengan pendekatan visualisasi menggunakan t-SNE, UMAP, dan PaCMAP secara bersamaan. Penelitian sebelumnya umumnya hanya mengandalkan pemodelan numerik tanpa visualisasi interaktif yang mendalam [11]. Dalam pemanfaatan LRFM dan K-Means saja, tanpa menggabungkan teknik visualisasi multidimensi akan ada keterbatasan dalam menganalisis hubungan kompleks dan sulit dalam pengambilan keputusan [10]. Padahal, eksplorasi visual dapat memberikan nilai tambah yang signifikan dalam menginterpretasi hasil segmentasi pelanggan [11]. Visualisasi multidimensi, seperti t-SNE, UMAP, dan PaCMAP membantu mengungkap pola tersembunyi dalam data berdimensi tinggi, memudahkan identifikasi segmen pelanggan dan mendukung pengambilan Keputusan yang efektif [12].

Penelitian ini bertujuan untuk menerapkan analisis LRFM dalam mengukur perilaku pelanggan dan memvisualisasikannya menggunakan tiga teknik proyeksi, yaitu t-SNE, UMAP, dan PaCMAP. Dengan pendekatan ini, pola kluster pelanggan dapat diidentifikasi dan dianalisis secara lebih informatif. Nilai kebaruan dari penelitian ini terletak pada penggabungan antara teknik *behavioral analytics* LRFM dan proyeksi visual berbasis *projection learning*. Pendekatan ini dapat dimanfaatkan oleh perusahaan, dalam hal ini sebuah perusahaan kursus musik di Surabaya dalam menyusun strategi retensi dan akuisisi pelanggan yang lebih terarah, sehingga berdampak langsung pada peningkatan profitabilitas. Penelitian ini juga berkontribusi dalam

pengembangan penerapan analitik pelanggan dalam bidang konsentrasi Manajemen Bisnis dan Kewirausahaan pada program studi Teknik Industri, khususnya dalam mendukung pengambilan keputusan berbasis data.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksploratif untuk melakukan segmentasi pelanggan berbasis model Length, Recency, Frequency, dan Monetary (LRFM), serta memvisualisasikan hasil segmentasi menggunakan teknik reduksi dimensi t-SNE, UMAP, dan PaCMAP. Prosedur penelitian ini terdiri dari beberapa tahap utama yang dijelaskan secara berurutan sebagai berikut.



Gambar 1. Diagram Alir Penelitian

Secara umum, desain penelitian ini melibatkan lima tahap utama: (1) input data sebagai akuisisi data historis transaksi pelanggan kursus, (2) perhitungan atribut LRFM, (3) pemberian skor dan segmentasi pelanggan, (4) reduksi dimensi dan visualisasi menggunakan t-SNE, UMAP, dan PaCMAP, serta (5) analisis hasil visualisasi dan profil pelanggan. Gambar 1 menyajikan alur lengkap proses penelitian. Penelitian ini didasarkan pada segmentasi pelanggan dengan menganalisis perilaku transaksi pelanggan. Data yang digunakan diperoleh dari aktivitas operasional sebuah lembaga kursus musik selama periode Juli 2019 hingga April 2025. Dalam kurun waktu tersebut, tercatat sebanyak 670 pelanggan aktif, yang terdiri dari siswa, orang tua, maupun wali. Total transaksi yang terekam mencapai sekitar 3.600 entri. Informasi utama dalam dataset meliputi nama pelanggan, tanggal transaksi, serta nominal transaksi dalam satuan Rupiah, sebagaimana dirinci pada Tabel 1.

Tabel 1. Data Mentah Penelitian

No	Atribut	Deskripsi
1	Tanggal	Tanggal transaksi pelanggan, tersebar mulai 2019 hingga 2025
2	Nama	Identitas pelanggan yang melakukan transaksi
3	Nominal	Jumlah uang yang dibayar pelanggan

Langkah pertama dalam proses segmentasi pelanggan adalah menghitung empat komponen utama dari model LRFM, yaitu Length, Recency, Frequency, dan Monetary. Nilai Length (L) dihitung sebagai selisih hari antara tanggal transaksi terakhir dan tanggal transaksi pertama yang dilakukan oleh masing-masing pelanggan [13]. Komponen Recency (R) diukur berdasarkan selisih hari antara tanggal acuan (*snapshot date*) dan tanggal transaksi terakhir pelanggan. Adapun Frequency (F) merepresentasikan jumlah total transaksi yang dilakukan oleh seorang pelanggan dalam periode pengamatan, sementara Monetary (M) mencerminkan total nilai pembelian pelanggan dalam satuan mata uang rupiah. Setelah seluruh nilai LRFM dihitung, masing-masing atribut tersebut kemudian diubah ke dalam bentuk skor dengan skala 1 sampai 5. Proses skoring ini dilakukan menggunakan metode pembagian kuantil (*quintile*), di mana data dibagi menjadi lima kelompok yang sama besar [14]. Pelanggan dengan nilai tertinggi pada suatu dimensi akan memperoleh skor 5, sedangkan yang terendah akan memperoleh skor 1. Sebagai contoh, pelanggan dengan nilai Recency terendah akan mendapat skor tertinggi karena menandakan transaksi paling baru [15]. Pemberian skor ini bertujuan untuk menyamakan skala antar dimensi dan mempermudah proses klasifikasi pelanggan ke dalam segmen tertentu. Hasil dari proses skoring ini kemudian dikombinasikan dalam analisis segmentasi pelanggan yang dijelaskan lebih lanjut pada subbab berikutnya.

```

If L ≥ 4 and R ≥ 4 and F ≥ 4 and M ≥ 4:
    Segment = "Best Customers"
Elif L ≥ 4 and (F ≥ 4 or M ≥ 4):
    Segment = "Loyal Customers"
Elif L ≤ 3 and R ≥ 3 and (F ≥ 3 or M ≥ 3):
    Segment = "Potential"
Else:
    Segment = "At Risk"

```

Gambar 2. Algoritma Penentuan Segmentasi Pelanggan

Setelah nilai dan skor dari setiap dimensi LRFM diperoleh, langkah selanjutnya adalah melakukan klasifikasi pelanggan ke dalam beberapa segmen strategis. Segmentasi ini dilakukan dengan menggabungkan nilai dari empat dimensi, yaitu *Length*, *Recency*, *Frequency*, dan *Monetary*, untuk membentuk karakteristik unik dari setiap pelanggan [16]. Penggunaan skor kuantil pada tahap sebelumnya memudahkan dalam pengelompokan karena telah menyederhanakan nilai numerik menjadi kategori ordinal. Pengelompokan pelanggan dilakukan dengan pendekatan berbasis aturan (*rule-based segmentation*), di mana setiap kombinasi skor memiliki arti tersendiri yang ditunjukkan oleh Gambar 2. Dalam penelitian ini, ditetapkan empat segmen utama, yaitu **Best Customers**, **Loyal Customers**, **Potential**, dan **At Risk**. Misalnya, pelanggan yang memiliki skor tinggi pada keempat dimensi digolongkan sebagai **Best Customers**, yang menunjukkan bahwa mereka telah lama berlangganan, masih aktif, sering membeli, dan memiliki nilai pembelian tinggi. Sebaliknya, pelanggan dengan skor rendah dalam berbagai dimensi dikategorikan sebagai **At Risk** karena menunjukkan tanda-tanda penurunan hubungan dengan bisnis. Pendekatan segmentasi ini bersifat deterministik dan mudah diterapkan oleh praktisi bisnis karena tidak memerlukan algoritma kompleks. Selain itu, metode ini transparan

dalam penentuan aturan segmentasi, sehingga dapat langsung diterjemahkan ke dalam strategi pemasaran yang spesifik. Hasil segmentasi selanjutnya menjadi dasar dalam eksplorasi visual menggunakan teknik reduksi dimensi, yang bertujuan untuk menggambarkan distribusi pelanggan secara intuitif.

Model LRFM menghasilkan empat dimensi perilaku pelanggan yang saling berkaitan. Namun, visualisasi langsung terhadap data berdimensi empat sulit dilakukan, terutama ketika jumlah pelanggan cukup banyak [17]. Oleh karena itu, dibutuhkan teknik *dimensionality reduction* atau reduksi dimensi untuk memproyeksikan data LRFM ke dalam dua dimensi sehingga dapat divisualisasikan dalam bentuk grafis dua dimensi yang mudah dipahami. Dalam penelitian ini digunakan tiga teknik reduksi dimensi yang umum dalam analisis data nonlinier, yaitu *t-distributed stochastic neighbor embedding (t-SNE)*, *uniform manifold approximation and projection (UMAP)*, dan *pairwise controlled manifold approximation (PaCMAP)*. Metode t-SNE unggul dalam mempertahankan struktur lokal antar data, UMAP menyeimbangkan antara struktur lokal dan global, sementara PaCMAP memiliki fleksibilitas lebih tinggi dalam menjaga distribusi kluster. Ketiga metode ini diterapkan secara terpisah terhadap dataset LRFM untuk membandingkan pola kluster yang dihasilkan. Hasil dari setiap teknik divisualisasikan dalam bentuk *scatter plot* dua dimensi, di mana setiap titik mewakili satu pelanggan, dan warna titik menunjukkan segmen pelanggan hasil klasifikasi sebelumnya. Visualisasi ini bertujuan untuk mengungkap pola klusterisasi yang mungkin tersembunyi di balik data numerik, sehingga membantu pemahaman strategis oleh pengambil keputusan non-teknis. Selain itu, visualisasi interaktif juga digunakan untuk mengeksplorasi lebih lanjut hubungan antara atribut LRFM dan perilaku pelanggan secara keseluruhan.

Seluruh tahapan dalam penelitian ini diimplementasikan menggunakan bahasa pemrograman Python versi 3.11, dengan lingkungan eksekusi berbasis cloud menggunakan Google Colaboratory. Pemilihan platform ini mempertimbangkan fleksibilitas akses, kemudahan integrasi dengan Google Drive untuk penyimpanan data, serta dukungan penuh terhadap pustaka analisis data dan visualisasi. Data input dalam format Excel dibaca menggunakan pustaka `'pandas'`, sedangkan pembersihan data dilakukan secara otomatis menggunakan fungsi `'dropna'` dan `'datetime parsing'`. Untuk perhitungan LRFM dan proses skoring kuantil, digunakan fungsi `'rank'` dan `'qcut'` dari pustaka `'pandas'`. Segmentasi dilakukan dengan pendekatan logika kondisional yang diimplementasikan menggunakan fungsi `'apply'` pada `DataFrame`. Reduksi dimensi dilakukan menggunakan pustaka `'scikit-learn'` untuk t-SNE, `'umap-learn'` untuk UMAP, dan `'pacmap'` untuk PaCMAP. Seluruh parameter pada setiap algoritma disesuaikan agar memberikan hasil visualisasi yang stabil dan representatif. Visualisasi hasil reduksi dimensi serta analisis deskriptif tambahan dilakukan menggunakan `'plotly'` untuk grafik interaktif, serta `'matplotlib'` dan `'seaborn'` untuk grafik statistik tambahan. Proses ini memungkinkan interaktivitas dalam pengamatan kluster, seperti memperbesar wilayah tertentu, memfilter segmen pelanggan, hingga mengeksport gambar dan tabel ke dalam format Excel dan PNG. Alat bantu ini memberikan kemudahan dalam mengevaluasi strategi segmentasi dan efektivitas visualisasi secara menyeluruh.

3. HASIL DAN PEMBAHASAN

Berdasarkan hasil skoring LRFM, seluruh pelanggan berhasil dikelompokkan ke dalam empat segmen utama: *Best Customers*, *Loyal Customers*, *Potential*, dan *At Risk*. Gambar 3 menunjukkan distribusi segmen berdasarkan jumlah pelanggan dan persentasenya. Sebagian besar pelanggan masuk dalam kategori At Risk yaitu sebesar 47,4% (310 pelanggan), yang menunjukkan rendahnya retensi dan nilai transaksi. Sebaliknya, proporsi *Best Customers* relatif kecil, yaitu 18,5% (121 pelanggan) namun memiliki kontribusi besar terhadap total nilai pembelian. Segmen *Best Customers* dicirikan oleh skor tinggi pada semua dimensi, menandakan pelanggan setia dengan aktivitas pembelian tinggi dan nilai transaksi besar. Sementara itu,

pelanggan pada segmen Potential memiliki kecenderungan untuk berkembang menjadi pelanggan loyal jika didukung dengan strategi pemasaran yang tepat, seperti pemberian insentif atau follow-up layanan.

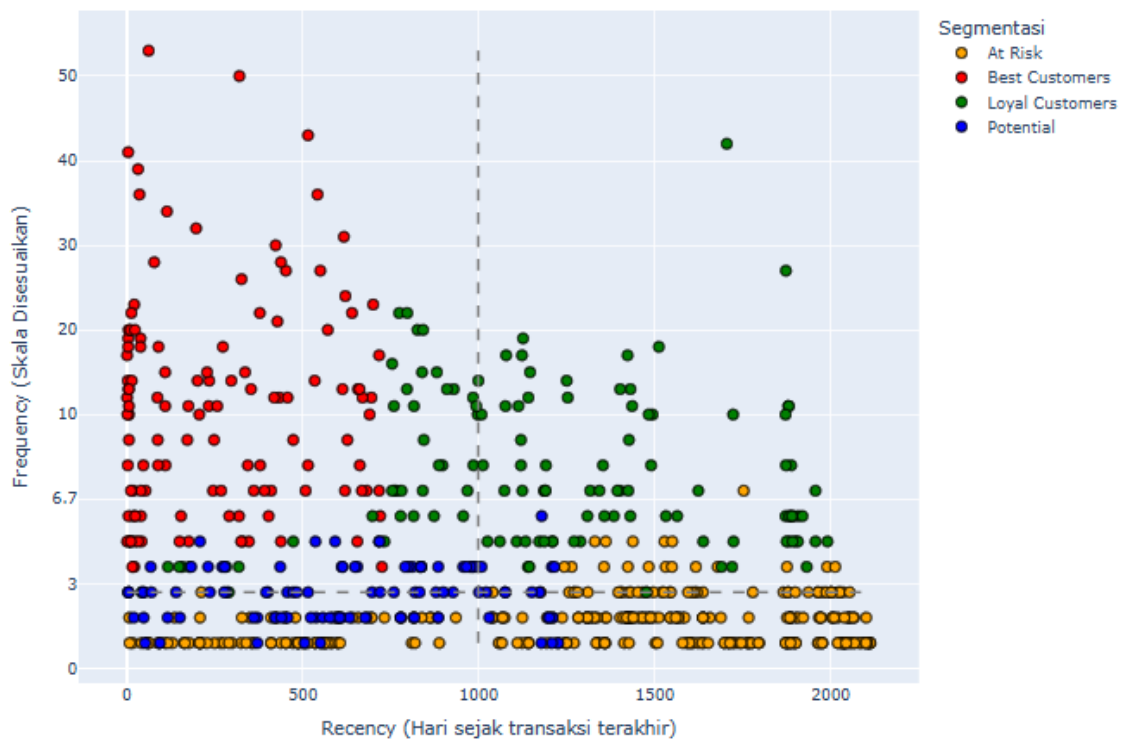


Gambar 3. Distribusi Pelanggan Tiap Segmen

Dalam visualisasi yang terbentuk pada Gambar 4, dapat dilihat bahwa kuadran 1 (bagian kanan atas) didominasi oleh warna hijau, yang merepresentasikan segmen *Best Customers*. Kuadran ini mencakup pelanggan dengan *recency* rendah (yang berarti mereka baru saja melakukan transaksi) dan *frequency* tinggi (sering bertransaksi). Ini adalah pelanggan paling aktif dan loyal saat ini. Mereka tidak hanya sering kembali, tetapi juga baru-baru ini berinteraksi, menjadikan mereka target utama untuk program loyalitas, penghargaan, dan retensi eksklusif. Kuadran 2 (bagian kiri atas) cenderung dipenuhi oleh warna merah, yang mengindikasikan segmen *Loyal Customers*. Pelanggan dalam kuadran ini memiliki *frequency* tinggi namun *recency* tinggi, yang artinya mereka adalah pelanggan yang dulunya sangat aktif, namun akhir-akhir ini sudah tidak melakukan transaksi lagi. Ini bisa menjadi indikasi adanya kehilangan minat atau pergeseran kebutuhan. Mereka penting untuk ditindaklanjuti dengan strategi *re-engagement* seperti penawaran khusus atau kampanye personalisasi.

Sementara itu, kuadran 3 (bagian kiri bawah) cenderung didominasi oleh warna kuning, yang mewakili segmen *At Risk*. Pelanggan di kuadran ini memiliki *frequency* rendah dan *recency* tinggi, menandakan mereka jarang melakukan transaksi dan terakhir kali bertransaksi pun sudah cukup lama. Mereka adalah kelompok yang sudah hampir tidak aktif, dan memiliki risiko tinggi untuk *churn*. Strategi retensi yang agresif atau reaktivasi khusus diperlukan di sini jika masih dianggap memiliki potensi nilai jangka panjang. Terakhir, kuadran 4 (bagian kanan bawah) memperlihatkan banyak warna biru, yang sebagian besar berasal dari segmen *Potential*. Mereka memiliki *recency* rendah (baru saja bertransaksi), tetapi *frequency* rendah (baru membeli satu atau dua kali). Meskipun belum aktif sepenuhnya, mereka menunjukkan tanda awal keterlibatan yang bisa berkembang. Ini adalah kelompok ideal untuk edukasi produk, *onboarding*, dan kampanye yang mendorong pembelian berulang agar bisa naik kelas menjadi pelanggan loyal.

Gambar 5 menunjukkan Treemap pertama yang memetakan 50 pelanggan teratas berdasarkan nilai *monetary*, di mana beberapa pelanggan menyumbang nilai transaksi yang sangat besar. Treemap kedua menampilkan kontribusi *monetary* dari setiap segmen pelanggan. Dari visual ini, terlihat bahwa segmen “*Best Customers*” mendominasi nilai penjualan, disusul “*Loyal Customers*”, sementara “*At Risk*” dan “*Potential*” menyumbang jauh lebih kecil. Ini menegaskan pentingnya mempertahankan pelanggan utama.



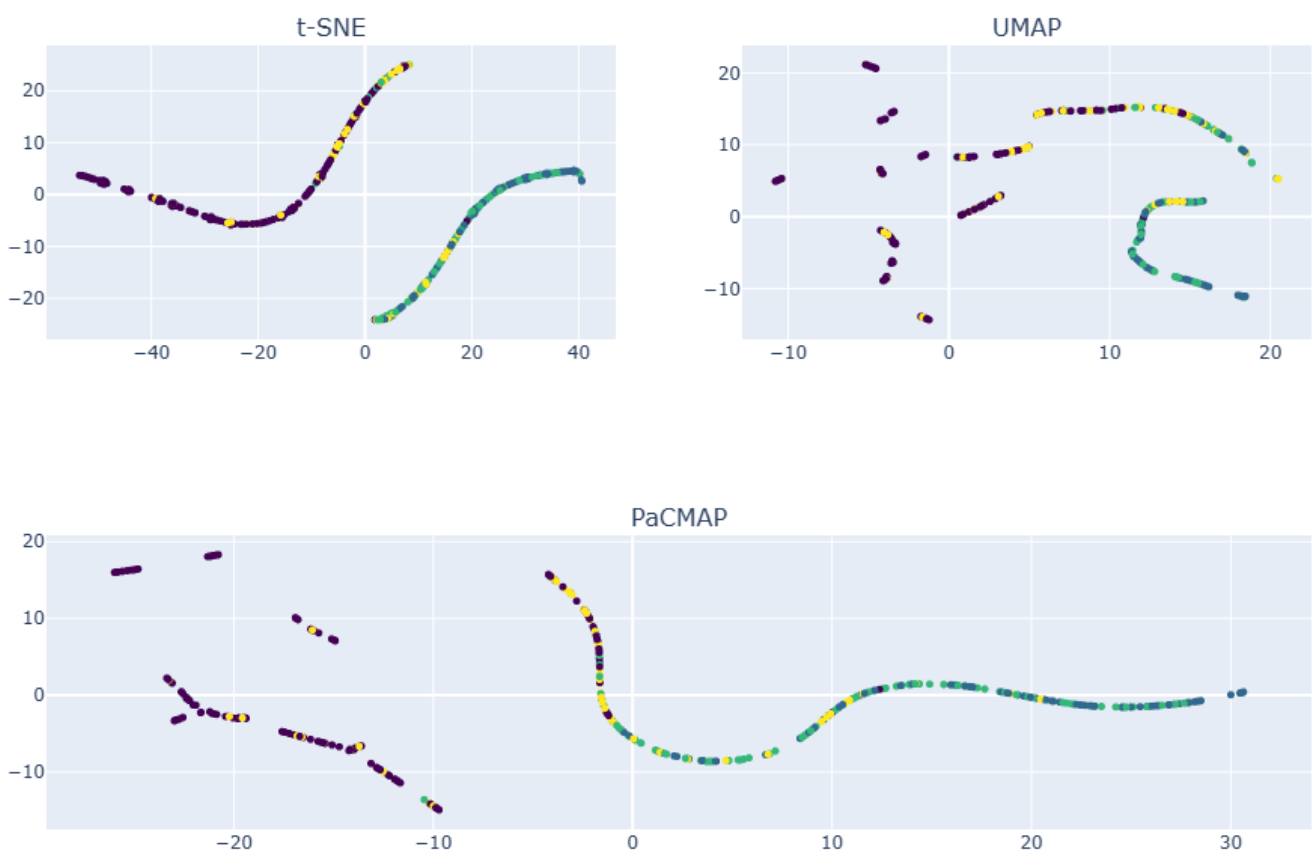
Gambar 4. Distribusi Pelanggan Tiap Segmen



Gambar 5. Peta Pohon Visualisasi Segmen Pelanggan

Visualisasi hasil segmentasi dilakukan menggunakan tiga teknik reduksi dimensi yang berbeda untuk memberikan sudut pandang yang beragam terhadap pola data. Gambar 6 menunjukkan hasil proyeksi dua dimensi menggunakan t-SNE, UMAP, dan PaCMAP. Warna pada tiap titik merepresentasikan segmen pelanggan. Visualisasi menggunakan *t-distributed Stochastic Neighbor Embedding* (t-SNE) menunjukkan penyebaran data pelanggan yang

membentuk pola lengkung berlapis yang cukup rapi. Kluster pelanggan dari berbagai segmen cenderung saling terpisah, menunjukkan bahwa t-SNE mampu menjaga hubungan jarak lokal dengan baik, yaitu pelanggan yang memiliki kesamaan skor LRFM berada dalam area yang berdekatan. Ini membuat t-SNE sangat efektif dalam mengidentifikasi kelompok pelanggan homogen, khususnya untuk mendeteksi kluster dengan karakteristik mirip seperti pelanggan dengan nilai *Length* dan *Monetary* tinggi. Hasil visualisasi dari *Uniform Manifold Approximation and Projection* (UMAP) menunjukkan distribusi yang lebih padat dan kompak di beberapa area, dengan adanya celah antar-kluster yang relatif jelas. UMAP unggul dalam menjaga struktur global maupun lokal dari data asli, sehingga pola transisi antar segmen pelanggan dapat diamati secara lebih menyeluruh. Pada grafik ini, kelompok pelanggan dari segmen *Loyal* dan *Potential* dapat dikenali melalui distribusi titik yang menyebar secara moderat namun tidak terlalu bertumpuk, menunjukkan ketepatan UMAP dalam menjaga relasi topologi antar dimensi LRFM. Visualisasi dengan *Pairwise Controlled Manifold Approximation Projection* (PaCMAP) menghasilkan distribusi yang lebih merata dengan pola melengkung halus. Salah satu keunggulan PaCMAP terlihat dari kemampuannya mempertahankan jarak antar titik dari segmen yang berbeda, tanpa kehilangan detail relasi internal dari setiap kluster. Bentuk garis berliku yang dihasilkan menunjukkan adanya gradasi antar pelanggan yang memungkinkan analisis lebih mendalam terhadap pelanggan yang berada di antara dua segmen, seperti peralihan dari *Potential* ke *Loyal*. PaCMAP juga mempertahankan keterbacaan.



Gambar 6. Plot Visualisasi Gabungan t-SNE, UMAP dan PaCMAP

Berdasarkan hasil visualisasi t-SNE, UMAP, dan PaCMAP, dapat dilakukan perbandingan kualitatif menggunakan tiga indikator: *separation clarity*, *cluster compactness*, dan *outlier detection*. Visualisasi t-SNE menunjukkan pemisahan segmen pelanggan yang sangat jelas, dengan kelompok “*Best Customers*” dan “*At Risk*” tampak terpisah secara ekstrem. Kluster

yang terbentuk cenderung kompak, meskipun beberapa mengalami distorsi bentuk karena karakteristik non-linear t-SNE. Selain itu, t-SNE efektif dalam mengidentifikasi pelanggan outlier yang menyendiri dari kelompok utama. Sementara itu, UMAP menghasilkan klaster yang sangat padat dan seimbang, dengan struktur visual yang mencerminkan hubungan alami antar data. Meskipun pemisahannya tidak sejelas t-SNE, UMAP tetap mampu menampilkan batas antarklaster secara halus. Kelebihan utama UMAP adalah kemampuannya menjaga kontinuitas dan bentuk manifold data asli, meskipun deteksi outlier sedikit berkurang. PaCMAP memberikan pendekatan berbeda dengan menampilkan pola transisi yang menggambarkan perubahan antar segmen secara bertahap. Klaster pada PaCMAP terlihat memanjang namun tetap memberikan gambaran alami tentang perubahan nilai LRFM pelanggan. Outlier pun dapat teridentifikasi dengan baik di area yang menjauh dari jalur utama. Dari ketiganya, UMAP dinilai paling seimbang untuk memahami struktur pelanggan secara menyeluruh, sedangkan t-SNE dan PaCMAP lebih cocok digunakan untuk eksplorasi visual awal dan deteksi pelanggan ekstrem. Ketiganya saling melengkapi dalam membangun strategi segmentasi berbasis visualisasi untuk meningkatkan profitabilitas bisnis.

Hasil visualisasi UMAP memberikan gambaran yang kuat terkait pola pengelompokan empat segmen pelanggan berdasarkan skor LRFM. Pada proyeksi dua dimensi UMAP, segmen *Best Customers* menempati posisi yang paling terpisah dan terkonsentrasi, menunjukkan bahwa pelanggan dalam kelompok ini memiliki karakteristik yang sangat konsisten, yakni nilai *Length*, *Recency*, *Frequency*, dan *Monetary* yang tinggi. Klaster ini tampak menonjol dibanding segmen lainnya dan mudah diidentifikasi dalam peta visual, menjadikannya sasaran strategis utama untuk dipertahankan dan ditingkatkan kontribusinya terhadap profitabilitas bisnis. Segmen *Loyal Customers* cenderung mengelompok di area yang berdekatan, namun dengan sebaran yang lebih longgar, mencerminkan pelanggan setia dengan keterlibatan tinggi tetapi mungkin belum mencapai nilai *Monetary* maksimum. Segmen *Potential* tersebar lebih lebar, menandakan variasi karakteristik dan potensi pertumbuhan yang beragam. Sementara itu, segmen *At Risk* cenderung menyendiri atau terletak di pinggir distribusi, menunjukkan rendahnya aktivitas dan kontribusi pelanggan di kelompok ini. Dengan kekuatan UMAP dalam mempertahankan struktur manifold dan hubungan lokal antar data, pola-pola ini dapat digunakan sebagai dasar visual yang kuat untuk mendukung strategi retensi pelanggan bernilai tinggi serta aktivasi ulang pelanggan yang berisiko.

Hasil visualisasi menggunakan t-SNE, UMAP, dan PaCMAP menunjukkan pola klaster yang konsisten, dengan segmentasi "*Best Customers*" dan "*At Risk*" yang terpisah jelas. Temuan ini sejalan dengan penelitian Santoso dan Hamdani [7], yang menggunakan K-Means untuk mengidentifikasi pelanggan bernilai tinggi, meskipun tanpa visualisasi dimensi rendah. Selain itu, struktur lokal yang kuat pada hasil t-SNE dan keseimbangan global dari UMAP juga mendukung temuan Wang *et al* [8], yang menekankan keunggulan masing-masing metode dalam representasi visual data pelanggan.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa model segmentasi pelanggan berbasis analisis *Length*, *Recency*, *Frequency*, dan *Monetary* (LRFM) yang dikombinasikan dengan metode visualisasi non-linear seperti t-SNE, UMAP, dan PaCMAP mampu memberikan gambaran yang lebih eksploratif dan mendalam terhadap perilaku pelanggan. Dari ketiga metode visualisasi yang digunakan, UMAP terbukti paling seimbang dalam hal kejelasan pemisahan segmen, kepadatan klaster, dan kemudahan interpretasi hasil. Visualisasi dua dimensi UMAP memperlihatkan bahwa segmen *Best Customers* berada dalam klaster yang sangat terpisah dan padat, mencerminkan stabilitas kontribusi pelanggan tersebut terhadap profitabilitas. Sementara itu, segmen lainnya seperti *Loyal*, *Potential*, dan *At Risk* juga dapat diidentifikasi secara visual, sehingga dapat digunakan sebagai dasar pengambilan keputusan dalam strategi pemasaran dan pengelolaan hubungan pelanggan. Pendekatan berbasis LRFM yang diperkuat dengan visualisasi interaktif

terbukti tidak hanya meningkatkan pemahaman terhadap struktur pelanggan, namun juga memberikan arah yang lebih strategis dalam peningkatan retensi dan optimasi nilai pelanggan. Rekomendasi lanjutan dari penelitian ini adalah menerapkan hasil segmentasi ke dalam sistem rekomendasi atau program loyalitas berbasis data. Selain itu, penelitian lanjutan dapat mengeksplorasi model klasifikasi otomatis terhadap segmen pelanggan dengan memasukkan variabel eksternal lainnya seperti demografi atau perilaku digital pelanggan.

DAFTAR PUSTAKA

- [1] A. C. Utomo, A. Handojo, dan T. Octavia, "Customer Loyalty Segmentation in Online Store Using LRFM and MLRFM in Combination with RM K-Means Algorithm". *Jurnal Teknik Informatika (JUTIF)*, vol. 5, no. 2, artikel 1497, 2024.
- [2] I. Sakti, A. Mareta, dan I. Wasito, "Fraud Detection in Mobile Phone Recharge Transactions Using K-Means and t-SNE Visualization," *Sinkron: Jurnal dan Penelitian Teknik Informatika*, vol. 9, no. 1, pp. 248–258, 2025.
- [3] T. Djatnika, "Manajemen Hubungan Pelanggan, Pengaruhnya terhadap Retensi Pelanggan pada Unit Usaha Koperasi Simpan-Pinjam (CRM, and Its Impact on Customer Retention at a Credit Union)," *Jurnal Administrasi Niaga, Politeknik Negeri Bandung (POLBAN)*, vol. 1, no. 1, pp. 1016–1025, 2023.
- [4] I. D. V. Sasikirana, A. S. Dewi, Q. A. Khayzuran, S. P. Firdausy, dan D. O. Radianto, "Strategi pemasaran digital yang efektif untuk meningkatkan daya saing perusahaan di era digital," *Profit: Jurnal Manajemen, Bisnis dan Akuntansi*, vol. 3, no. 2, hlm. 166–177, 2024.
- [5] D. Qiao, X. Ma, dan J. Fan, "Federated t-SNE and UMAP for Distributed Data Visualization," dipublikasikan Desember 2024.
- [6] H. M. T. Ramadhan, A. Pertiwi, G. A. Defanka, dan A. P. Sari, "Clustering RFM (Recency, Frequency, Monetary) publisher gim menggunakan algoritma K-Means," *Seminar Nasional Informatika Bela Negara (SANTIKA)*, vol. 3, 2023.
- [7] B. Shen, "E-commerce customer segmentation via unsupervised machine learning," dalam *Proc. 2nd Int. Conf. Comput. Data Sci.*, pp. 1–7, Jan. 2021.
- [8] C. Nisak dan E. Sugiharti, "Customer Lifetime Value Clustering Using K-Means Algorithm with Length Recency Frequency Monetary Model to Enhance Customer Relationship Management," *Journal of Advances in Information Systems and Technology*, vol. 6, no. 1, pp. 383–391, 2024.
- [9] M. Husnah dan R. A. Vinarti, "Customer Segmentation Analysis Using LRFM Based Product and Brand Dimensions," dalam *2023 2nd Int. Conf. for Innovation in Technology (INOCON)*, 2023.
- [10] Y. Wang, H. Huang, C. Rudin, dan Y. Shaposhnik, "Understanding how dimension reduction tools work: An empirical approach to deciphering t-SNE, UMAP, TriMap, and PaCMAP for data visualization," *Journal of Machine Learning Research*, vol. 22, pp. 1–73, Jul. 2021.
- [11] A. Salmanian, M. A. Raji, B. O. Womack, dan H. D. Pfister, "DimVis: Interpreting Visual Clusters in Dimensionality Reduction With Explainable Boosting Machine," 2024.
- [12] N. U. Jannah, N. Y. Setiawan, dan W. Purnomo, "Segmentasi pelanggan B2B dengan model LRFM menggunakan algoritma K-Means," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 1, 2017.
- [13] D. Qiao, X. Ma, dan J. Fan, "Federated UMAP for Distributed Data Visualization," dalam Dong et al., *Fed-UMAP*, 2024.
- [14] E. Santoso dan A. U. Hamdani, "Segmentasi pelanggan pada PT. TAB menggunakan algoritma K-Means clustering," *JIKI (Jurnal Ilmu Komputer dan Informatika)*, vol. 4, no. 2, pp. 135–145, Des. 2023.

- [15] M. Yousuff, R. Babu, dan A. Rathinam, “Nonlinear dimensionality reduction based visualization of single-cell RNA sequencing data,” *Journal of Analytical Science and Technology*, 2024.
- [16] H. J. Joosse, dkk., “Haematology dimension reduction, a large scale application to regular care haematology data”, *BMC Medical Informatics and Decision Making*, 2025.
- [17] H. J. Wilbert, A. F. Hoppe, A. Sartori, S. F. Stefenon, dan L. A. Silva, “Recency, Frequency, Monetary value, Clustering, and Internal and External Indices for Customer Segmentation from Retail Data”, *Algorithms*, vol. 16, no. 9, art. 396, 2023.