IT-EXPLORE
Jurnal Penerapan Teknologi Informasi dan Komunikasi
ISSN 2828-7940, e-ISSN 2829-1727, DOI: 10.24246/itexplore.v5i1.2026.pp75-85
Volume 05 Nomor 01 Tahun 2026
Journal homepage: https://ejournal.uksw.edu/itexplore

Segmentasi pelanggan pada penjualan perusahaan Aair minum XYZ dengan analisis RFM dan K-Means clustering

Geizka Wasito Adi [1], Alz Danny Wowor [2]
[1,2] Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Salatiga, Jawa Tengah, Indonesia

Informasi Artikel

Riwayat Artikel:

Dikirim:
21 Mei 2025
Direvisi:
20 Juni 2025
Diterima:
22 Juni 2025

Kata Kunci:

Algoritma K-Means,
Clustering,
Segmentasi,
Strategi Pemasaran,
Pelanggan.

ABSTRAK

Karena banyaknya persaingan di industri retail air minum, strategi pemasaran yang tepat dan terfokus diperlukan. Salah satu metode yang digunakan adalah segmentasi pelanggan dengan analisis RFM (*Recency*, *Frequency*, and *Monetary*). Selain itu, metode klasterisasi berbasis pembelajaran mesin, clustering K-Means, juga digunakan. Tujuan dari penelitian ini adalah untuk menggambarkan hasil penelitian dalam bentuk grafik yang dapat digunakan untuk menganalisis pola pelanggan berdasarkan karakteristiknya. Penelitian ini juga menggunakan metode *clustering* untuk menguji tiga metrik, dan hasilnya menunjukkan bahwa indeks *calinski harabaz* adalah metrik yang sesuai untuk melanjutkan ke tahap segmentasi dengan nilai 10286. Untuk setiap evaluasi klaster, terbentuk klaster optimal di mana indeks calinski harabaz terbentuk menjadi tiga klaster yang lebih masuk akal untuk disegmentasi. Ini berbeda dengan metrik evaluasi lainnya yang hanya menghasilkan dua klaster; misalnya, ketika menyebarluaskan data transaksi penjualan air minum XYZ, ditemukan bahwa *diamond* memiliki 1061 pelanggan, *star diamond* memiliki 2005 pelanggan, dan *woodsale* memiliki 422 pelanggan dari total peljualan air minum. Diharapkan hasil penelitian ini dapat membantu manajemen perusahaan air minum XYZ dan bidang pemasaran lainnya menghadapi persaingan yang semakin ketat.

Keywords:

K-Means Algorithm,
Clustering,
Segmentation,
Marketing Strategy,
Customers.

ABSTRACT

A suitable and targeted marketing plan is required because of the intense competition in the retail drinking water sector. Customer segmentation using RFM (Recency, Frequency, and Monetary) analysis is one of the techniques employed. Additionally, K-Means clustering, a clustering technique based on machine learning, is employed. This study's goal is to present the findings in the form of graphs that can be used to examine consumer trends according to their attributes. With a value of 10286, the Calinski Harabaz index is a suitable metric to move on to the segmentation step in this study, which also tests three metrics using the clustering method. An ideal cluster is created for every cluster evaluation by dividing the Calinski Harabaz index into three more manageable clusters. This contrasts with other evaluation metrics that only yield two clusters. For instance, when XYZ drinking water sales transaction data was distributed, it was discovered that, out of the total drinking water sales, woodsale had 422 customers, diamond had 1061 customers, and star diamond had 2005 customers. The management of the XYZ drinking water company and other marketing fields are expected to encounter more intense competition as a result of the study's findings.

This is an open access article under the CC BY-SA license

Corresponding Author:

Geizka Wasito Adi
Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana
Jl. Diponegoro No.52-60, Sidorejo, Salatiga, Jawa Tengah, Indonesia
Email: 672021207@student.uksw.edu

1. PENDAHULUAN

Dalam dunia modern, dinamika bisnis ditandai dengan tingkat persaingan yang semakin intens. Ini terutama berlaku dalam industri retail air minum kemasan, yang terus berkembang karena munculnya berbagai merk baru. Kondisi ini menuntut setiap bisnis, termasuk perusahaan air minum XYZ, untuk membuat strategi pemasaran yang fleksibel dan efisien untuk meningkatkan profitabilitas dan mempertahankan pansa pasar. Memahami target pasar potensial sangat penting untuk menentukan strategi pemasaran yang tepat. Dalam situasi seperti ini, segmentasi pelanggan menjadi metode penting yang memungkinkan bisnis untuk membagi basis pelanggannya menjadi kelompok yang homogen berdasarkan fitur tertentu. Dengan demikian, bisnis dapat menyesuaikan produk, layanan, dan pesan pemasarannya secara lebih khusus dan relevan untuk setiap segmen pasar. Ini adalah langkah penting mengingat kompleksitas bidang bisnis dan pemasaran saat ini.

Terlepas dari fakta bahwa segmentasi pelanggan telah banyak dibahas dan digunakan, masalah utama yang sering dihadapi adalah bagaimana melakukan segmentasi yang akurat, terukur, dan dapat diterapkan secara praktis, terutama untuk data transaksi yang besar dan kompleks. Model analisis RFM (Recency, Frequency, and Monetary) telah dikenal luas sebagai alat yang efektif untuk memahami perilaku pembelian pelanggan dan menargetkan strategi pemasaran. Tiga variabel utama digunakan dalam model ini: frekuensi (seberapa baru pelanggan bertransaksi), frekuensi (seberapa sering pelanggan bertransaksi), dan monetary (berapa banyak uang yang dihabiskan pelanggan) [1]. Namun, untuk menggunakan RFM pada kumpulan data transaksi massif, diperlukan teknik lanjutan untuk menemukan pola tersembunyi dan otomatis mengelompokkan pelanggan [2].

Penelitian – penelitian sebelumnya banyak yang memanfaatkan Teknik *clustering*, seperti K-Means, untuk segmentasi pelanggan [3]. Namun, evaluasi klaster yang dihasilkan seringkali tidak menunjukkan metrik mana yang paling tepat dan memberikan interpretasi yang paling logis untuk tujuan segmentasi. Penelitian ini sangat penting karena akan secara eksplisit mengisi celah tersebut dengan menggunakan analisis RFM dan algoritma k-means clustering serta melakukan pengujian komparatif terhadap beberapa metrik evaluasi klaster (seperti skor silhouette, indeks Davies Bouldin, dan indeks Calinski Harabasz). Tujuan dari penelitian ini adalah untuk menemukan metrik yang paling sesuai untuk menghasilkan jumlah klaster optimal dan paling rasional untuk segmentasi pelanggan perusahaan air minum XYZ. Penelitian ini akan berkonsentrasi pada indeks calinksi harabaz, yang telah terbukti mampu menghasilkan tiga klaster yang lebih masuk akal untuk ditafsirkan dibandingkan dengan metrik lainnya. Selain itu, penelitian ini akan menampilkan hasil segmentasi dalam bentuk visualisasi grafik yang jelas dan mudah dipahami, yang akan menggambarkan profil dan karakteristik unik setiap segmen pelanggan seperti bintang safir, safir, perak, silver, bronze, dan woodsale. Penelitian ini juga akan mengeksplorasi potensi [4]. Dengan demikian, hasil penelitian ini diharapkan dapat memberikan informasi yang lebih terukur dan aplikatif bagi manajemen Perusahaan air minum XYZ dalam merumuskan strategi pemasaran yang lebih terfokus, meuingkatkan loyalitas pelanggan, mengoptimalkan pengelolaan inventaris dan akhirnya menghadapi persaingan pasar yang semakin ketat secara lebih efektif [5].

2. KAJIAN PUSTAKA

2.1. State of The Art

Studi tentang segmentasi pelanggan telah banyak dilakukan, terutama dengan Teknik Clustering. Berbagai penelitian telah menerapkan pendekatan ini pada berbagai industri untuk mempelajari perilaku pelanggan dan membuat strategi pemasaran yang lebih efisien [6]. Penelitian oleh Dinata et al., yang menggunakan analisis kelompok k-means pada data sepeda motor, merupakan contohnya. Dengan menggunakan analisis pengukuran performa untuk akurasi, recall, dan ketepatan, tujuan penelitian adalah untuk menentukan sepeda motor yang ideal. Ini menunjukkan bahwa aplikasi clustering dapat digunakan untuk menentukan

preferensi produk.

Penelitian lain yang relevan dilakukan oleh Savitri dkk [7]. Dengan tujuan yang sama untuk melakukan segmentasi pelanggan pada klinik kecantikan. Hasil penelitian tersebut menunjukan bahwa pembagian pelanggan menjadi dua klaster adalah yang paling optimal, dan setelah menemukan klaster optimal, dilakukan pengelompokkan pelanggan pada hasil dari Teknik *clustering*. Studi ini menekankan pada identifikasi jumlah klaster optimal dan pengelompokkan pelanggan berdasarkan hasil *clustering*.

Selanjutnya, Cristover, Toba, dan Suteja [8] juga melakukan segmentasi nasabah pada perusahaan perbankan mandiri cabang Puri Indah. Penelitian ini bertujuan untuk mengetahui kelompok nasabah dan melakukan prediksi yang dapat digunakan sebagai bahan *follow-up* perusahaan terhadap nasabah dengan menggunakan hasil prediksi dari nilai RFM. Hasil yang didapatkan adalah dua klaster, dimana klaster satu merupakan segmen nasabah dengan dominan kelompok segmen *low value* dan *medium value customer*, sementara klaster kedua memiliki dominan kelompok segmen *top customer*. Penelitian ini secara spesifik menggabungkan RFM dengan *clustering* untuk tujuan *follow-up* dan identifikasi nilai nasabah.

Selain itu, Tampubolon, Butar Butar, dan Rochimah [9] juga melakukan segmentasi pelanggan majalah pada situs web *e-commerce* menggunakan K-Means++ dan metode RFM. Penelitian ini menunjukan bagaimana kombinasi metode tersebut dapat digunakan untuk mengelompokkan pelanggan berdasarkan perilaku pembelian mereka di *platform* digital. Hal ini relevan karena menunjukan aplikasi RFM dan K-Means dalam konteks penjualan digital. Penelitian oleh oleh Yusak dkk [10]. Juga mengkaji analisis segmentasi pelanggan ritel *online* menggunakan *k-means clustering* berdasarkan model *recency, frequency, monetary* (RFM). Studi ini semakin memperkuat penggunakan kombinasi RFM dan K-Means sebagai pendekatan yang kuat untuk segmentasi pelanggan dalam lingkungan ritel.

Meskipun penelitian-penelitian terdahulu telah berhasil menerapkan *clustering*, khususnya K-Means dengan atau tanpa model RFM, untuk segmentasi pelanggan di berbagai sektor, terdapat beberapa aspek yang dapat diperdalam. Penelitian oleh Savitri dkk, dan Cristover dkk. Sama-sama menghasilkan dua klaster optimal, menunjukan bahwa hasil jumlah klaster optimal dapat bervariasi tergantung pada dataset dan metrik evaluasi. Hal ini mengindikasi adalah celah dalam konsistensi dan interpretasi jumlah klaster optimal paling relevan untuk segmentasi strategis.

2.2. Landasan Teori

A. Model RFM

Model RFM (*recency, frequency,* dan *monetary*) merupakan pendekatan yang telah terbukti efektif dan banyak digunakan dalam *customer relationship management* (CRM) dan pemasaran. Konsep ini pertama kali diperkenalkan oleh Arthur Hughes pada tahun 1994 [11] dan sejak itu menjadi standar industry untuk menganalisis nilai pelanggan dan memprediksi perilaku pembelian di masa depan. Model RFM mengklasifikasikan pelanggan berdasarkan tiga dimensi utama dari riwayat transaksi mereka. *Recency*, mengukur seberapa baru transaksi terakhir pelanggan, pelanggan yang baru saja melakukan pembelian kemungkinan lebih tinggi untuk merespon tawaran baru dibandingkan mereka yang sudah lama tidak bertransaksi. Skor *recency* yang rendah menunjukan pelanggan yang aktif, sementara skor tinggi menandakan pelanggan yang lama tidak melakukan transaksi. *Frequency*, mengukur seberapa sering pelanggan melakukan pembelian dalam periode waktu tertentu. Pelanggan dengan frekuensi pembelian tinggi cenderung lebih loyal dan sering berinteraksi dengan produk atau layanan perusahaan. *Monetary*, mengukur total jumlah uang yang dihabiskan pelanggan dalam periode waktu tertentu. Pelanggan dengan nilai *monetary* tinggi adalah mereka yang memberikan kontribusi pendapatan tersebar bagi perusahaan. Keuntungan utama modal RFM terletak pada kemampuannya untuk mengidentifikasi pelanggan bernilai tinggi secara objektif berdasarkan data transaksi yang terukur. Dengan mengkombinasikan ketiga metrik ini, perusahaan dapat perusahaan dapat mengelompokkan pelanggan ke dalam segmen-segmen yang berbeda, memungkinkan personalisasi strategi pemasaran dan alokasi sumber daya yang lebih efesie. Pendekatan ini sering digunakan untuk segmentasi pasar, mengidentifikasi pelanggan yang berharga, dan memprediksi respons terhadap promosi [12].

B. Segmentasi Pelanggan

Segmentasi pelanggan adalah proses pembagian basis pelanggan menjadi kelompok-kelompok yang lebih kecil dan homogen, dimana setiap kelompok memiliki karakteristik dan kebutuhan yang serupa. Ini adalah konsep inti dalam pemasaran hubungan dan merupakan langkah awal yang krusial dalam perumusan strategi bisnis yang efektif. Tujuan utama segmentasi adalah untuk mengingkatkan pemahaman perusahaan

terhadap pelanggannya dan memungkinkan penyesuaian produk, layanan, serta pesan pemasaran agar lebih relevan dan menarik bagi setiap segmen. Menurut Kotler dan Keller (2016) [13], segmentasi pasar memungkinkan perusahaan untuk melayani konsumennya secara lebih efektif, sehingga meningkatkan kepuasan pelanggan dan profitabilitas. Dengan mengetahui segmen pelanggan, perusahaan dapat membuat strategi pemasaran yang lebih terarah.

C. *K-Means Clustering*

K-Means *clustering* adalah salah satu algoritma *clustering* non-hierarki yang paling popular dan banyak digunakan dalam *data mining*. Tujuannya adalah untuk mengelompokkan *n* objek data ke dalam klaster, dimana setiap objek termasuk dalam klaster dengan pusat (*centroid*) terdekatnya [14]. Algoritma ini secara iteratif untuk meminimalkan jumlah kuadrat jarak antara objek data dan pusat klaster masing-masing. K-means dikenal karena kesederhanaan dan efesiensi komputasinya, menjadikannta pilihan ideal untuk dataset yang besar.

Menurut Han, Kamber, & Pei (2012) Langkah – langkah umum algoritma k-means adalah sebagai berikut [15] :

1. Inisialisasi *Centroid*, memiliki *k* objek secara acak sebagai pusat klaster awal.

2. Penetapan Klaster, setiap objek data ditugaskan ke klaster yang *centroid*-nya paling dekat. Jarak eucledian sering digunakan sebagai metrik kemiripan.

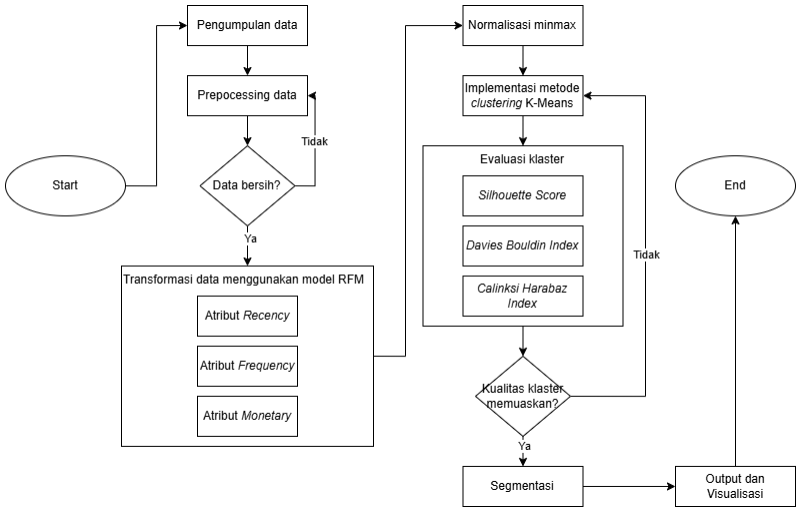
3. Pembaruan *Centroid*, *centroid* baru dihitung sebagai rata-rata dari semua objek data yang ditugaskan ke klaster tersebut.

4. Iterasi, langkah 2 dan 3 diulang hingga *centroid* tidak lagi berubah secara signifikan atau jumlah iterasi maksimum tercapai.

K-means berbeda dengan metode klasifikasi karena bersifat *unsupervised learning*, yaitu mengelompokkan data tanpa adanya label kategori yang telah ditentukan sebelumnya. Hal ini sangat berguna dalam analisis data transaksi yang sangat banyak dan kompleks, karena dapat secara alami mengidentifikasi kelompok-kelompok pelanggan berdasarkan kemiripan perilaku mereka.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan eksperimental untuk menentukan kelompok pelanggan dengan model RFM dan teknik *clustering k-mean*. Analisis ini mencangkup beberapa tahapan, mulai dari pengumpulan data, *preprocessing data*, *agregation for RFM* (transformasi data), *clustering k-means with evaluation matrix* (Implementasi model), segmentasi, output dan visualisasi . Semua tahapan ini dapat dilihat pada Gambar 1.



Gambar 1: Alur penelitian

Alur penelitian dimulai dari pengumpulan data yang dibutuhkan untuk penelitian ini. *Dataset* yang digunakan dalam penelitian ini adalah data transaksi penjualan produk air minum XYZ dengan total 16433 row dalam periode bulan April tahun 2024 yang ada pada perusahaan air minum XYZ. Berikut cuplikan *dataset* dapat dilihat pada Tabel 1.

Tabel 1 Cuplikan *dataset*

CustomerId	CustomerName	DoFDate	ItemAmount
7696	Endra – SO	17 April 2024	Rp.8800000
7451	Hartini – Toko	15 April 2024	Rp.170000
4740	Kevin – Toko	17 April 2024	Rp.255000
7854	Kilakilo – Toko	17 April 2024	Rp.170000
13999	Toko Sembako MDRA	17 April 2024	Rp.204000

Setelah selesai menentukan *dataset* yang akan digunakan dalam penelitian, sebelum data digunakan data diolah terlebih dahulu pada tahap *preprocessing data* untuk dilakukan pengecekan agar hasil analisis yang didapatkan sesuai dengan keinginan. Tahap *preprocessing data* adalah proses dimana data akan diakukan *cleaning* atau pembersihan terlebih dahulu, dari data duplikat, *missing data* atau data hilang dan meminimalisirnya outlier untuk meningkatkan kualitas data agar mudah untuk dilakukan analisis [16].

Dalam proses transformasi data, skor *frequency* dihitung dengan membandingkan transaksi pelanggan. Skor *Frequency* ditentukan dari total pelanggan melakukan transaksi, dan skor *Monetar* ditentukan dari perhitungan jumlah total pelanggan setiap melakukan transaksi. Hasil transformasi data dapat dilihat pada Tabel 2.

Tabel 2 Data hasil transformasi RFM

CustomerId	*Recency*	*Frequency*	*Monetary*
9	1	2	Rp.85000
20	6	3	Rp.1032500
22	6	9	Rp.5032000
29	1	3	Rp.166500
36	7	1	Rp.188000

Setelah hasil didapatkan, pada tahap selanjutnya adalah implementasi metode klaster. Sebelum itu skor hasil tersebut akan dinormalisasikan terlebih dahulu agar data yang akan digunakan dapat terhindari dari potensi bias dan distorsi yang disebabkan oleh skala fitur yang berbeda dalam data, dan memudahkan dalam mengorganisasian data. Perhitungan dengan normalisasi minmax dapat dilihat pada Pers. (1).

$$X_{norm} = \frac{x' - min(x)}{max(x) - min(x)} \left(new_{max(x)} - new_{min(x)}\right) + new_{min(x)} \quad (1)$$

Hasil normalisasi telah didapatkan barulah dilakukan *clustering* menggunakan metode K-Means. Metode *clustering* memiliki kinerja dalam mengelompokkan dasar item ke dalam bentuk fitur baru pada grup *k*. Kemudian menetapkan titik data ke pusat klaster terdekat, titik- titik data yang dekat dengan pusat *k* akan terbentuk sebuah klaster baru [17]. Kombinasi algoritma K-Means dengan evaluasi klaster dalam menentukan klaster yang paling optimal pada suatu dataset dengan cara menjalankan K-Means dengan berbagai jumlah klaster yang berbeda, untuk mengukur kualitas pengelompokan pada setiap jumlah klaster akan menentukan skor evaluasi terbaik

Scoring menggunakan persamaan rumus kuartil merupakan metode pemberian skor kepada pelanggan berdasarkan nilai *recency, frequency* dan *monetary*. dimana data untuk setiap metrik diurutkan dan dibagi menjadi empat kelompok sama besar menggunakan nilai – nilai kuartil, lalu setiap kelompok diberi skor dengan skala ordinal yaitu satu sampai empat untuk mempresentasikan tingkatan nilai metriks tersebut [18], dengan skor yang lebih tinggi menunjukan perilaku pelanggan yang lebih menguntungkan, kecuali nilai tinggi pada artibut *recency* menandakan bahwa pelanggan telah melakukan transaksi paling terbaru bahkan sebaliknya jika nilai *recency* rendah menandakan bahwa pelanggan lama tidak melakukan transaksi. Perhitungan dengan rumus kuartil *frequency* dan *monetary* menggunakan Pers (2) dan rumus kuartil untuk

recency dapat juga dilihat pada Pers (3).

$$Score_{F\ dan\ M} = \begin{cases} 4 \ \ jika\ x > Q1_{F\ dan\ M} \\ 3 \ \ jika\ x \leq Q1_{F\ dan\ M} \ dan\ x > Q2_{F\ dan\ M} \\ 2 \ \ jika\ x \leq Q1_{F\ dan\ M} \ dan\ x \leq Q2_{F\ dan\ M} \ dan\ x > Q3_{F\ dan\ M} \\ 1 \ \ jika\ x \leq Q1_{F\ dan\ M} \ dan\ x \leq Q2_{F\ dan\ M} \ dan\ x \leq Q3_{F\ dan\ M} \end{cases} \quad (2)$$

$$Score_R = \begin{cases} 4 & jika\ x & \leq Q1_R \\ 3 & jikaQ1_R < & x \leq Q2_R \\ 2 & jikaQ2_R < & x \leq Q3_R \\ 1 & jika\ x & > Q3_R \end{cases} \quad (3)$$

Hasil *scoring* menggunakan skala kuartil telah ditetapkan, tahap terakhir adalah menentukan kelompok pelanggan berdasarkan ketentuan nilai karakteristiknya. Untuk jenis kelompok pelanggan bersarkan nilai karakteristiknya akan dibagi menjadi enam, yaitu *star diamond, diamond, gold, silver, bronze,* dan *woodsale*, untuk menentukan jenis kelompok pelanggan hasil skor akan mencari nilai karakteristik yang sesuai dengan jenis kelompoknya, sebagaimana yang tertera pada Tabel 3.

Tabel 3 Kelompok pelanggan segmentasi

Kelompok Pelanggan	Karakteristiknya	*Scoring*
Star Diamond	Pelanggan terbaik Paling baru, sering berbelanja, dan menghabiskan paling banyak. Sangat loyal.	12>
Diamond	Pelanggan yang sangat baik. Baru atau sering berbelanja dan/atau menghabiskan banyak. Potensi loyalitas tinggi.	10-11
Gold	Pelanggan yang baik. Cukup baru atau sering berbelanja atau menghabiskan cukup banyak.	8-9
Silver	Pelanggan dengan nilai RFM sedang. Mungkin sudah lama tidak berbelanja atau frekuensi/nilai transaksinya tidak terlalu tinggi.	6-7
Bronze	Pelanggan dengan nilai RFM rendah. Sudah lama tidak berbelanja dan frekuensi/nilai transaksinya rendah. Berisiko untuk hilang.	4-5
Woodsale	Pelanggan dengan skor sangat rendah di semua metrik. Kemungkinan pelanggan baru dengan transaksi awal yang kecil atau pelanggan yang hampir hilang.	<3

4. HASIL DAN PEMBAHASAN

4.1. *Preprocessing data*

Pada tahap *preprocessing*, pembersihan data dilakukan dengan menghapus baris yang kosong, duplikat dan nilai null. Setelah data kosong, dan data duplikat dihapus, proses *prepocessing* dimulai dengan menghapus kolom selain kolom *customerid, dofdate,* dan *itemamount*, sehingga dataset keseluruhan yang valid berjumlah 3488. Tidak ada kolom lain yang digunakan karena hanya kolom ini yang diperlukan untuk penerapan model RFM (*recency, frequency, monetary*). Untuk membedakan setiap pelanggan, kolom *customerid* berfungsi sebagai indentifikasi unik, yang memungkinkan pengelompokan transaksi berdasarkan individu, kolom *dofdate* penting digunakan untuk menghitung *frequency* (jumlah transaksi yang dilakukan) dan interval waktu sejak transaksi terakhir. Namun ,data *monetary* (jumlah uang yang dikeluarkan pelanggan) diberikan oleh kolom *itemamount*, yang merupakan bagian penting lain dari analisis RFM. Penggunaan data ini secara eksklusif memastikan bahwa focus pada perilaku pembelian pelanggan seperti yang ditangkap oleh model RFM, yang merupakan dasar dari penelitian ini untuk melakukan segmentasi pelanggan. Hasil dari langkah ini ditunjukan dalam Tabel 4.

Tabel 4 Hasil *prepocessing*

Kolom	Keterangan
CustomerId	ID Pelanggan.
DoFDate	Tanggal pelanggan melakukan transaksi.
ItemAmount	Jumlah uang yang dikeluarkan saat melakukan transaksi.

4.2. Transformasi data

Kemudian, model RFM digunakan untuk menganalisis data transaksi yang tersedia. Data dikelompokkan kedalam kolom *recency, frequency,* dan *monetary*. Kolom *recency* mengandung jumlah hari terakhir dimana pelanggan melakukan transaksi dan kolom *frequency* mengandung jumlah kali pelanggan melakukan transaksi. Hasil *preprocessing* yang akan diolah dengan model RFM ditunjukan pada Tabel 2.

4.3. Normalisasi data

Setelah melalui tahap transformasi dengan model RFM, data akan dilakukan normalisasi agar skala data tidak jauh. Normalisasi MinMax disebut juga *scaling* Min-Max merupakan Teknik normalisasi data untuk menskalakan nilai-nilai fitur numberik dalam dataset ke dalam rentan tertentu, antara 0 dan 1. Tabel 5 adalah hasil dari normalisasi.

Tabel 5 Data normalisasi

CustomerId	*Recency*	*Frequency*	*Monetary*
9	0.04	0.015152	0.000206
20	0.24	0.030303	0.003082
22	0.24	0.121212	0.015221
29	0.04	0.030303	0.000454
36	0.28	0.000000	0.000519

4.4. Implementasi metode *clustering k-means*

Implementasi metode *clustering k-means* menggunakan tiga matrik evaluasi dalam pengujian ini. *Silhouette score* (SS), *Davies bouldin index* (DBI), dan *Calinski harabaz index* (CHI), pada *silhouette score*, nilai yang mendekati satu menandakan pembentukan klaster yang padat dan terpisah dengan baik. *Davies bouldin index* memiliki nilai rendah ataupun yang paling rendah, yang mengidentifikasikan pembentukan klaster yang kompak dan terpisah dengan baik. Sementara, *Calinski harabaz index* dengan nilai tertinggi menunjukan klaster yang padat serta terpisah dengan baik.

Hasil pengujian menggunakan 10 klaster menunjukan bahwa nilai *silhouette score* adalah 0,5865 menghasilkan klaster optimal pada dua klaster. Sedangkan nilai *davies bouldin index* adalah 0,5378 menghasilkan klaster optimal pada dua klaster. Nilai *calinski harabaz index* adalah 10286.50 menghasilkan klaster optimal pada klaster tiga, dapat dilihat pada Tabel 6. Meskipun ketiga metrik menunjukan hasil klaster yang baik, indeks *calinski harabaz index* dinilai lebih masuk akal daripada metrik lainnya untuk

segmentasi, karena memberikan nilai optimal yang lebih tinggi, mengarahkan pada pembentukan tiga klaster yang dianggap lebih relevan dan dapat diinterpretasikan dibandingkan metrik lain yang mungkin hanya menghasilkan dua klaster optimal.

Tabel 6 Hasil pengujian ketiga metrik evaluasi

Clustering	*Silhouette Score*	*Davies Blouldin Index*	*Calinski Harabaz Index*
K-Means	0.5865	0.5378	10286.50

4.5. Segmentasi

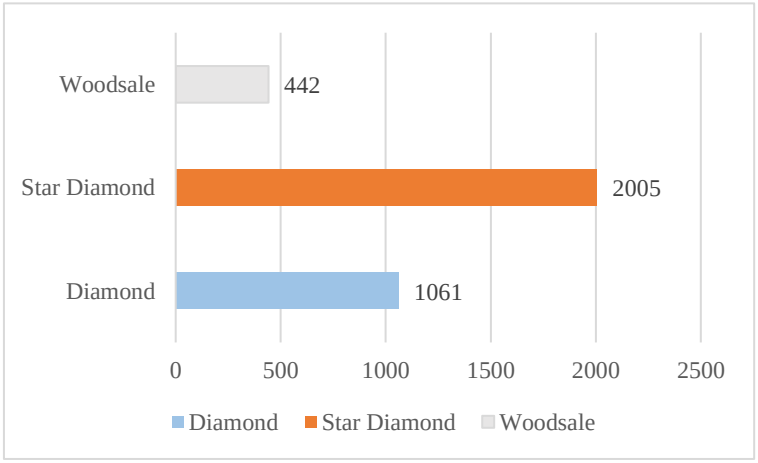
Pada tahap segmentasi, algoritma k-means mengelompokkan pelanggan berdasarkan karakteristiknya mereka. Jumlah klaster yang terbentuk adalah tiga (klaster 0, klaster 1, klaster 2) karena hasil pengujian metrik evaluasi *clustering*, khususnya *calinski harabaz index* (CHI), menunjukan bahwa nilai optimal terbentuk pada tiga klaster. Indeks *calinski harabaz* mencapai nilai tertinggi 10286.50 pada pembentukan tiga klaster, yang dianggap lebih masik akal untuk segmentasi dibandingkan metrik evaluasi lainnya yang hanya menghasilkan dua klaster. Meskipun Tabel 3 mengidentifikasi enam kelompok pelanggan potensial (*Star diamond, Diamond, Gold, Silver, Bronze, Woodsale*), hasil optimal dari *algoritma k-means* dengan data spesifik ini dan metrik evaluasi yang dipilih (CHI) secara empiris mengarah pada pembentukan tiga klaster yang paling representatif. Tabel 7 merangkum hasil pengelompokkan *cluster* berdasarkan karakteristiknya.

Tabel 7 Hasil pengelompokan *cluster*

Cluster	R_Skor	F_Skor	M_Skor	RFM_Skor	Kelompok Segmen
Cluster 0	3	4	4	11	*Diamond*
Cluster 1	4	4	4	12	*Star Diamond*
Cluster 2	1	1	1	3	*Woodsale*

4.6. Output dan visualisasi

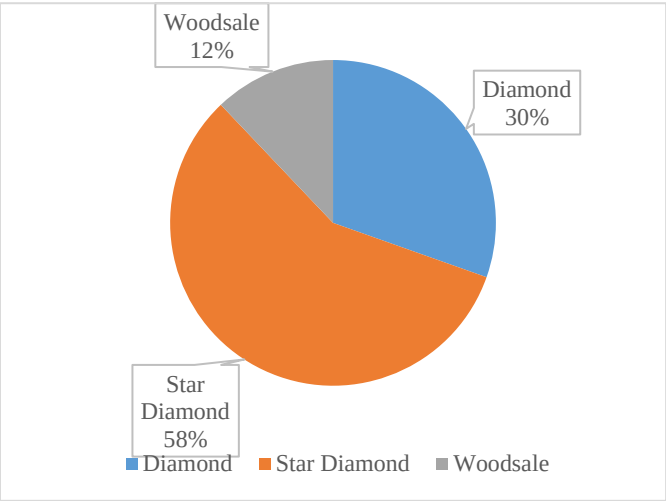
Dalam Upaya untuk meningkatkan penjualan produk perusahaan air minum XYZ, penelitian ini bertujuan untuk mengidentifikasi jenis kelompok pelanggan berdasarkan ciri-ciri mereka yang mempengaruhi proses pengambilan Keputusan manajemen. Hasil ini dicapai dengan bantuan alat berbasis Phyton [19], dimana *Google Colaboratory* dan *Microsoft Excel* berfungsi sebagai fondasi untuk format dataset [20]. Gambar 2 menunjukan grafik jumlah pelanggan tiap klaster berdasarkan kelompok segmennya, dan Gambar 3 menunjukan presentase jumlah pelanggan berdasarkan grafik pie chart.



Gambar 2: Jumlah pelanggan berdasarkan kelompoknya

Pada Gambar 2 grafik batang menunjukan jumlah pelanggan aktual dalam setiap klaster segmentasi yang dihasilkan oleh algoritma k-means. *Star diamond* merupakan segmen terbesar dengan 2005 pelanggan. Kelompok ini diidentifikasikan sebagai pelanggan yang paling setia, dengan skor RFM tertinggi yaitu 12, menandakan mereka adalah pelanggan yang paling baru bertransaksi, paling sering berbelanja, dan menghabiskan uang paling banyak. *Diamond* merupakan segmen menengah dengan 1061 pelanggan.

Pelanggan di segmen ini memiliki skor RFM antara 10-11, menunjukan mereka adalah pelanggan yang sangat baik, baru, atau sering berbelanja dan menghabiskan banyak uang, serta memiliki potensi loyalitas tertinggi. *Woodsale* merupakan segmen terkecil dengan 422 pelanggan. Kelompok ini ditandai dengan skor RFM yang sangat rendah yaitu tiga kebawah, mengindikasikan kemungkinan pelanggan baru dengan transaksi awal yang kecil atau pelanggan yang hampir hilang.



Gambar 3: Presentase kelompok pelanggan berdasarkan *cluster* yang terbentuk

Grafik pie pada Gambar 3 menggambarkan proporsi relatif dari setiap segmen pelanggan terhadap total keseluruhan pelanggan yang disegmentasi. Kelompok segmen *Star diamond*, menyumbang 58% dari total pelanggan, proporsi ini menegaskan dominasi segmen pelanggan paling bernilai dan setia, yang telah diidentifikasikan sebagai pelanggan "*Star Diamond*". *Diamond*, mencangkup 30% dari total pelanggan menunjukan vahwa segmen pelanggan yang sangat baik, dengan potensi loyalitas tinggi, merupakan bagian yang signifikan dari basis pelanggan. *Woodsale* merupakan 12% dari total pelanggan, meskipun presentasinya paling kecil, segmen ini tetap penting untuk diidentifikasikan guna merancang strategi retensi atau aktivasi kembali.

V. SIMPULAN DAN SARAN

Penelitian ini bertujuan untuk mengatasi tantangan persaingan ketat di industri retail air minum, khususnya bagi perusahaan air minum XYZ, yang menuntut strategi pemasaran yang lebih terfokus. Masalah utama yang diidentifikasikan adalah kebutuhan akan segmentasi pelanggan yang akurat, terukur dan dapat diimplementasikan secara praktis, terutama dalam menghadapi data transaksi yang besar dan kompleks.

Penelitian ini mengadopsi pendekatan segmentasi pelanggan dengan memadukan analisis RFM (*recency, frequency,* dan *monetary*) dan algortima k-means *clustering*. Analisis RFM digunakan untuk memahami perilaku pembelian pelanggan berdasarkan seberapa baru mereka bertransaksi (*recency*), seberapa sering mereka bertransaksi (*frequency*), dan seberapa banyak uang yang mereka habiskan (*monetary*). Selanjutnya, algoritma K-Means *clustering* diterapkan pada skor RFM yang di normalisasikan untuk mengelompokkan pelanggan ke dalam segmen-segmen homogen secara otomatis.

Dengan demikian, penelitian ini berhasil mengidentifikasikan dan mengelompokkan pelanggan perusahaan air minum XYZ ke dalam tiga segmen utama berdasarkan perilaku pembelian mereka. Pertama adalah *star diamond*, merupakan segmen terbesar dengan 2005 pelanggan, diindetifikasikan sebagai pelanggan paling setia dengan skor RFM tertinggi (12), yang berarti mereka adalah pelanggan yang paling baru tertransaksi, paling sering berbelanja, dan menghabiskan uang paling banyak. Segmen ini menyumbang 58% dari total pelanggan. Kedua adalah *diamond*, merupakan segmen menengah dengan 1061 pelanggan, memiliki skor RFM antara 10-11, menunjukan mereka adalah pelanggan yang sangat unik, baru, atau sering berbelanja dan menghabiskan banyak uang, serta memiliki potensi loyalitas tertinggi. Segmen ini mencangkup 30% dari total pelanggan. Sedangkan yang ketika adalah *woodsale*, yang merupakan segmen terkecil dengan 422 pelanggan, ditandai dengan skor RFM yang sangat rendah, dibawah tiga. Mengindikasikan kemungkinan pelanggan baru dengan transaksi awal yang kecil atau pelanggan yang hampir hilang. Segmen

ini menyumbang 12% dari total pelanggan.

Setiap segmen pelanggan memiliki fitur yang berbeda dan unik, dan hasil dari penelitian ini dapat memberikan informasi yang terukur dan aplikatif bagi manajemen perusahaan air minum XYZ. Dengan pemahaman mendalam tentang karakteristik setiap segmen , perusahaan dapat menyesuaikan produk, layanan, dan strategi pemasaran secara personal dan relevan. Hal ini diharapkan dapat membantu perusahaan menghadapi persaingan yang semakin ketat, mempertahankan loyalitas pelanggan, mengoptimalkan pengelolaan inventaris, dan pada akhirnya meningkatkan profitabilitas. Secara keseluruhan, penerapan analisis data dengan pendekatan IT ini memberikan contoh bagaimana pembelajaran mesin dapat di terapkan untuk menghasilkan informasi yang bermanfaat dalam bidang pemasaran dan bisnis.

DAFTAR PUSTAKA

[1] B. T. Kristanti, A. Junaidi, and E. P. Mandyartha, " Implementasi K-Means Clustering dalam Segmentasi Pelanggan Berdasarkan Usia, Pendapatan, dan Model Rfm (Studi Kasus: Lantikya Store Jombang)," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, Aug. 2024, doi: 10.23960/jitet.v12i3.4677.

[2] UNISBANK, "*Menghadapi Persaingan Dalam Dunia Bisnis Yang Semakin Ketat*." [online] available on https://www.unisbank.ac.id/v3/menghadapi-persaingan-dalam-dunia-bisnis-yang-semakin-ketat/

[3] R. K. Dinata, S. Safwandi, N. Hasdyna, and N. Azizah, "Analisis K-Means Clustering pada Data Sepeda Motor," *INFORMAL: Informatics Journal*, vol. 5, no. 1, p. 10, Apr. 2020, doi: 10.19184/isj.v5i1.17071.

[4] A. Widiyanto and A. Witanti, "Segmentasi Pelanggan Berdasarkan Analisis RFM Menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran (Studi Kasus PT Coversuper Indonesia Global)," *Jurnal Konstelasi*, vol. Vol. 1, no. Vol. 1 No. 1 (2021):Juni 2021, 2021.

[5] B. Basri, W. Gata, and R. Risnandar, "Analisis Loyalitas Pelanggan Berbasis Model Recency, Frequency, dan Monetary (RFM) dan Decision Tree pada PT. Solo," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 7, no. 5, pp. 943–950, Oct. 2020, doi: 10.25126/jtiik.2020752284.

[6] P. I. Pangestu, I. Hermanto, and D. Irmayanti, "Analisis Segmentasi Pelanggan Berbasis Recency Frequency Monetary (Rfm) Menggunakan Algoritma K-Means", *Jurnal Mahasiswa Teknik Informatika*, Vol. 7 No. 3 (2023).

[7] A. D. Savitri, F. Abdurrachman Bachtiar, and N. Y. Setiawan, "Segmentasi Pelanggan Menggunakan Metode K-Means Clustering Berdasarkan Model RFM Pada Klinik Kecantikan (Studi Kasus : Belle Crown Malang)," 2018. [Online]. Available: http://j-ptiik.ub.ac.id

[8] R. Cristover, H. Toba, and B. R. Suteja, "Segmentation and Formation of Customer Regression Model Based on Recency, Frequency and Monetary Analysis," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 2, Aug. 2022, doi: 10.28932/jutisi.v8i2.5075.

[9] A. L. M. Tampubolon, T. M. E. Y. Butar Butar, and S. Rochimah, "Segmentasi Pelanggan Majalah pada Situs Web E-Commerce dengan K-Means++ dan Metode RFM," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 6, pp. 1243–1252, Dec. 2024, doi: 10.25126/jtiik.2024118208.

[10] A. Yusak, N. Rumapea, D. Pratiwi, and S. Sari, "Analisis Segmentasi Pelanggan Ritel Online Menggunakan K-Means Clustering Berdasarkan Model Recency, Frequency, Monetary (RFM)," *Jurnal Sains dan Teknologi*, vol. 6, no. 3, pp. 292–299, 2024, doi: 10.55338/saintek.v6i3.4607.

[11] J.-T. Wei, S.-Y. Lin, and H.-H. Wu, "A review of the application of RFM model," *African Journal of Business Management*, vol. 4, no. 19, pp. 4199–4206, 2010, [Online]. Available: http://www.academicjournals.org/AJBM

[12] G. ASLANTAŞ, M. GENÇGÜL, M. RUMELLİ, M. ÖZSARAÇ, and G. BAKIRLI, "Customer Segmentation Using K-Means Clustering Algorithm and RFM Model," *Deu Muhendislik Fakultesi Fen ve Muhendislik*, vol. 25, no. 74, pp. 491–503, May 2023, doi: 10.21205/deufmd.2023257418.

[13] Kotler K, *"A Framework for Marketing Management.",* Pearson Publishing, 2022.

[14] D. Jollyta, "Segmentasi Pelanggan Hotel Labersa Untuk Mengetahui Tingkat Loyalitas Pelanggan Menggunakan Algoritma K-Means Dan RFM (Recency, Frequency, Monetary)," *Jurnal Mahasiswa Aplikasi Teknologi Komputer dan Informasi*, vol. 4, no. 3, pp. 135–139, 2022.

[15] S. Monalisa, "Klusterisasi Customer Lifetime Value dengan Model LRFM menggunakan Algoritma K-Means," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 2, pp. 247–252, May 2018, doi: 10.25126/jtiik.201852690.

[16] A. Wibowo, A. R. Handoko, and P. Korespondensi, "Segmentasi pelanggan ritel produk farmasi obat menggunakan metode data mining klasterisasi dengan analisis recency frequency monetary (RFM) termodifikasi," vol. 7, no. 3, 2020, doi: 10.25126/jtiik.202072925.

[17] D. A. Imanuel and G. Alfian, "Visualisasi Segmentasi Pelanggan Berdasarkan Atribut RFM Menggunakan Algoritma K-Means Untuk Memahami Karakteristik Pelanggan pada Toko Retail Online," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 12, no. 2, pp. 283–292, Apr. 2025, doi: 10.25126/jtiik.2025128619.

[18] A. Karami and Y. Widharto, "Perancangan Business Intelligence dan Segmentasi Pelanggan Menggunakan K Means Clustering Berdasarkan Rfm Model.", *Industrial Engineering Online Journal*, vol. 12, no. 1, Jan. 2023.

[19] N. H. Harani, C. Prianto, and F. A. Nugraha, "Segmentasi Pelanggan Produk Digital Service Indihome Menggunakan Algoritma K-Means Berbasis Python," *Jurnal Manajemen Informatika (JAMIKA)*, doi: 10.34010/jamika.v10i2.

[20] G. Sanjaya, ; Ade, R. Rinaldi, ; Fadhil, M. Basysyar, and ; Fathurrohman, "PENERAPAN ALGORITMA K-MEANS UNTUK MENGELOMPOKKAN BERBAGAI JENIS SMARTPHONE BERDASARKAN SPESIFIKASI." [Online]. Available: https://ikmi.ac.id/page/18/?lang=de