Coding : Jurnal Komputer dan Aplikasi
Volume 11, No. 03 (2023), hal 325-337

p-ISSN : 2338-493X
e-ISSN : 2809-574X

SISTEM SEGMENTASI PELANGGAN POTENSIAL MENGGUNAKAN METODE K-MEANS DAN ANALISIS RFM

[1]Diotivano, [2]Ikhwan Ruslianto, [3]Dian Prawira
[1][3]Jurusan Sistem Informasi, [2]Jurusan Rekayasa Sistem Komputer,
Fakultas MIPA Universitas Tanjungpura
Jalan Prof. Dr. H. Hadari Nawawi, Pontianak
Telp./Fax : (0561) 577963
e-mail: [1]diotivano@student.untan.ac.id, [2]ikhwanruslianto@siskom.untan.ac.id,
[3]dianprawira@sisfo.untan.ac.id

Abstrak

Pelanggan adalah potensi yang paling utama yang dimiliki setiap perusahaan. Maka dari itu, saat ini banyaknya perusahaan yang tidak hanya memfokuskan kegiatan pemasarannya pada pencarian pelanggan baru, tetapi lebih kepada upaya mempertahankan dan meningkatkan loyalitas pelanggan lama. Segmentasi pelanggan mengelompokkan pelanggan yang memiliki kesamaan karakteristik menjadi satu kelompok yang sama untuk membantu dalam proses penyesuaian produk dan layanan sehingga perusahaan dapat mengetahui pelanggan mana yang harus dipertahankan. Penelitian ini bertujuan untuk membuat sistem berbasis web yang berfungsi untuk melakukan segmentasi pelanggan potensial. Sistem Berbasis Web dibuat untuk memudahkan Perusahaan dalam melakukan segmentasi pelanggan dari sisi antarmuka. Metode clustering menggunakan algoritma K-Means clustering digunakan dalam melakukan segmentasi pelanggan. Analisis RFM digunakan untuk pemilihan variabel yang digunakan dalam melakukan clustering. RFM menyimpan informasi tentang waktu pembelian terbaru (recency), berapa kali pelanggan melakukan pembelian (frequency), dan rata-rata total uang yang dikeluarkan (monetary). Hasil dari clustering akan ditampilkan dalam bentuk visualisasi yang membantu penyajian data dan informasi sehingga lebih mudah dipahami dalam membaca hasil segmentasi pelanggan. Hasil dari penelitian ini berupa sistem segmentasi pelanggan potensial yang telah dilakukan pengujian fungsional, memperoleh hasil sesuai dengan rancangan fungsional sistem. Sedangkan pengujian interface kepada masyarakat umum memperoleh predikat sangat baik dengan persentase 87,59%.

Kata kunci: *Clustering*, *K-Means*, *RFM*, Segmentasi Pelanggan

1. PENDAHULUAN

1.1. Latar Belakang

Persaingan di dalam dunia bisnis dan pemasaran saat ini menjadi sangat ketat, maka dari itu perusahaan didorong untuk meningkatkan kemampuan dan keunggulan mereka dengan segala potensi yang ada. Salah satu potensi yang paling penting yang dimiliki oleh perusahaan adalah pelanggan, oleh karena itu pada saat ini banyaknya perusahaan yang lebih memfokuskan aktivitas pemasaran pada mempertahankan dan meningkatkan loyalitas pelanggan lama dibandingkan kepada usaha untuk pencarian pelanggan baru, dikarenakan biaya yang dikeluarkan untuk mendapatkan pelanggan baru dapat lebih banyak dibandingkan biaya mempertahankan pelanggan lama [1]. Dalam dunia bisnis, loyyalitas pelanggan merupakan kemauan pelanggan untuk terus membeli dan menggunakan suatu produk atau jasa dalam jangka waktu yang panjang [2]. Menjalin hubungan dengan pelanggan merupakan suatu strategi perusahaan dalam mempertahankan dan menumbuhkan basis pelanggan yang loyal. Maka perusahaan harus sadar bahwa menjalin hubungan dengan pelanggan salah satu hal yang penting untuk menunjang perkembangan dan kelangsungan perusahaan. Divisi pemasaran perusahaan harus mampu mencari cara untuk mengenali pelanggan potensialnya sehingga perusahaan dapat menerapkan strateginya untuk mempertahankan loyalitas pelanggannya. Segmentasi pelanggan merupakan sebuah praktik membagi pelanggan ke dalam

p-ISSN : 2338-493X
e-ISSN : 2809-574X

kelompok yang memiliki karakteristik serupa. Segmentasi pelanggan bermanfaat untuk mengidentifikasi pelanggan yang paling menguntungkan sehingga perusahaan dapat menyesuaikan produk dan layanan untuk memenuhi kebutuhan spesifik mereka [3].

PT. SID INTI DINAMIK adalah sebuah perusahaan di bidang produk pertanian. Fokus perusahaan adalah mendistribusikan produk pertanian di daerah Kalimantan Barat yang berupa pestisida, benih dan pupuk. Perusahaan ini tidak menjual produknya kepada konsumen akhir secara langsung menjualnya tetapi melalui retailer. Maka retailer yang berperan sebagai pelanggan pada perusahaan ini mempunyai peran penting dalam membantu perusahaan dalam memasarkan produk mereka kepada konsumen akhir. Mempertahankan pelanggan atau retailer merupakan hal yang sangat penting bagi perusahaan, tetapi perusahaan ini belum menerapkan suatu cara untuk menganalisis perilaku pelanggan atau dalam melakukan transaksi, sehingga belum diterapkannya strategi pemasaran yang tepat dalam mempertahankan pelanggan. Saat ini perusahaan masih memberikan strategi yang sama kepada seluruh pelanggan, seperti memberikan promosi yang serupa kepada setiap pelanggan dan memberikan reward atau bonus kepada pelanggan secara subjektif.

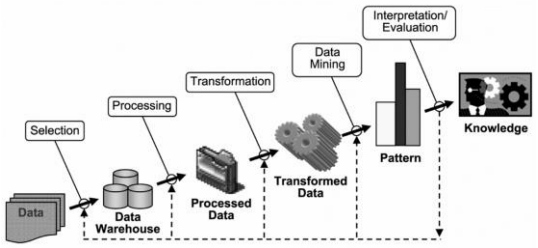
Permasalahan yang dihadapi oleh PT. SID INTI DINAMIK dalam mengatur strategi untuk mengelolah hubungan pelanggan dapat diatasi dengan dilakukannya segmentasi pelanggan potensial yaitu melakukan penggalian data menggunakan data transaksi penjualan pada periode tertentu. Segmentasi pelanggan mengelompokkan pelanggan berdasarkan karakter mereka masing-masing berdasarkan karakteristik pelanggan dalam melakukan transaksi. Pola karakteristik pembelian pelanggan dapat dideskripsikan melalui analisis RFM yaitu *Retency, Frequency,* dan *Monetary*. Analisis RFM (*Retency, Frequency,* dan *Monetary*) digunakan untuk pemilihan variabel yang digunakan dalam melakukan clustering. Analisis RFM menyimpan informasi tentang waktu pembelian terakhir (*recency*), berapa kali pelanggan melakukan transaksi (f*requency*), dan berapa total yang dihabiskan pelanggan dalam melakukan pembelian (*monetary*). Metode pengelompokan yang digunakan dalam melakukan segmentasi pelanggan potensial yaitu menggunakan algoritma *K-means clustering*. Berdasarkan dari semua permasalahan yang ada maka dari itu dengan adanya sistem segmentasi pelanggan potensial dan visualisasi berbasis web akan mempermudah perusahaan dalam mengenali pelanggan potensialnya dan memberikan kontribusi terhadap divisi pemasaran PT. SID INTI DINAMIK dalam merencanakan strategi dalam pengelolaan hubungan pelanggan sesuai dengan perilaku pembelian pelanggan.

2. LANDASAN TEORI

2.1. *Data Mining*

Data mining atau penggalian data merupakan proses yang menggunakan teknik statistika, matematika, *machine learning*, hingga memanfaatkan teknologi kecerdasan buata untuk mengekstrak dan mengumpulkan pengetahuan dan informasi yang berguna terkait dari database yang besar [4].

Data mining merupakan bagian utama dari *knowledge discovery in database* (KDD) yang merupakan proses yang membutuhkan beberapa langkah untuk mengubah data mentah menjadi informasi yang dapat digunakan (Pang-Ning Tan, dkk 2014). Proses ini terdiri dari serangkaian langkah seperti pada gambar 1.



Gambar 1. Proses Dari KKD (Knowledge Discovery in Database) [5]

Berikut ini adalah tahapan dari KKD berdasarkan gambar 1 [5]:

1. *Data Selection*
   Data Selection merupakan tahap pertama yang dilakukan sebelum dilakukannya proses penggalian informasi dalam KKD. Hasil dari tahap seleksi atau pemilihan data akan digunakan untuk proses penggalian informasi dimana data akan disimpan dalam suatu tempat yang terpisah dari *database* operasional utama.
2. *Pre-processing /cleaning*

p-ISSN : 2338-493X
e-ISSN : 2809-574X

Pre-processing atau *cleaning* merupakan tahap yang dilakukan untuk mengubah data yang telah diseleksi dalam format yang berguna dan efisien. Proses *pre-processing* data mencangkup seperti membuang data yang duplikat, memeriksa data yang inkonsisten, dan membetulkan kesalahan pada data yang telah dipilih.

3. *Transformation*
   Transformation merupakan tahap merubah format atau struktur data sehingga data tersebut sesuai digunakan untuk proses analisis dan *data mining*. Tahap *transformation* dalam KDD merupakan proses kreatif serta sangat bergantung pada tipe atau pola data yang akan dicari dalam *database*.
4. *Data Mining*
   Data Mining merupakan proses menggali pola atau informasi berharga dari kumpulan data yang besar dengan memakai metode tertentu ataupun tata cara tertentu. Teknik, metode, ataupun algoritma dalam data mining sangat beragam. Penentuan metode ataupun algoritma *data mining* yang tepat sangat tergantung pada tujuan serta proses KDD pada umumnya.
5. *Interpretation /evaluation*
   Interpretation atau *evaluation* adalah tahap menampilkan pola-pola atau informasi yang ditemukan dari proses *data mining* dalam bentuk yang valid, berguna, dan mudah dipahami. Tahap ini meliputi pengecekan apakah pola-pola atau informasi yang ditemukan tersebut berlawanan dengan fakta atau hipotesis yang ada sebelumnya.

2.2. Clustering

Clustering merupakan metode pengelompokan data yang banyak digunakan sebagai salah satu teknik *data mining*. C*lustering* bertujuan untuk mengelompokkan sejumlah populasi atau titik data ke dalam beberapa kelompok sehingga dalam setiap kelompok akan memiliki titik data atau karakteristik yang semirip mungkin. *Clustering* melakukan pengelompokkan pada data berdasarkan kemiripan antar objek, maka dari itu *clustering* termasuk sebagai metode *unsupervised learning* yang berarti bahwa pengelompokan yang dilakukan tidak mengacu pada pelabelan kelas dan pelatihan [6].

2.3. K-Means Clustering

K-means clustering merupakan algoritma analisis *cluster* non hirarki yang berusaha untuk mendapatkan kelompok data dengan memaksimalkan kesamaan karakteristik *cluster*, sehingga kelompok data dengan karakteristik serupa akan dikelompokan ke dalam satu *cluster* yang sama dan kelompok data dengan karakteristik yang berbeda akan dikelompokan kedalam *cluster* yang lain [7].

Berikut ini langkah-langkah *K-means clustering* [8]:

1. Menentukan nilai k sebagai jumlah cluster yang diinginkan.
2. Tetapkan nilai k secara acak sebagai lokasi pusat cluster awal.
3. Untuk setiap data, temukan pusat cluster terdekat. Untuk menentukan jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak *Euclidean* menggunakan persamaan berikut [9]:

$$D(i,j) = \sqrt{\left(X_{1i} - X_{1j}\right)^2 + \left(X_{2i} - X_{2j}\right)^2 + \cdots + \left(X_{ki} - X_{kj}\right)^2} \quad (1)$$

   Dimana:
   $D(i,j)$ = Jarak data ke *i* ke pusat cluster *j*
   X_{ki} = Data ke *i* pada atribut data ke *k*
   X_{kj} = Titik pusat ke *j* pada atribut ke *k*
4. Setelah mendapatkan C1, C2, … , Ck. Pada setiap cluster k, temukan pusat cluster baru, kemudian lakukan iterasi atau pembaruan lokasi setiap pusat cluster ke nilai baru dari pusat cluster. Untuk menentukan pusat cluster baru dapat menggunakan persamaan berikut:

$$\left(\sum \frac{a_i}{n}, \sum \frac{b_i}{n}, \ldots, \sum \frac{x_i}{n},\right) \quad (2)$$

   Dimana:
   a_i = total penjumlahan data cluster ke *i*
   n = jumlah data pada cluster *i*
5. Lakukan kembali langkah ke-3 hingga langkah terakhir hingga nilai dari titik centroid tidak berubah lagi.

p-ISSN : 2338-493X
e-ISSN : 2809-574X

2.4. Kuartil /*quartile*

Di dalam ilmu statistik, pengertian kuartil adalah titik, skor atau nilai yang membagi sekumpulan data yang terurut ke dalam empat bagian nilai data yang sama besar, yaitu masing masing bagian sebesar ¼ N. Maka akan ada tiga kuartil, yaitu kuartil pertama (Q1), kuartil kedua (Q2) atau median, dan kuartil ketiga (Q3). Ketiga kuartil ini akan membagi semua data yang terurut dari data yang telah selidiki menjadi empat bagian yang sama besar, masing-masing sebesar ¼ N. Berikut adalah rumus perhitungan untuk mencari posisi Q1, Q2, Q3 dalam baris data [10]:

$$Q_1 = \frac{1}{4}(n+1) \quad (3)$$

$$Q_2 = \frac{1}{2}(n+1) \quad (4)$$

$$Q_3 = \frac{3}{4}(n+1) \quad (5)$$

2.5. Metode Elbow

Metode *Elbow* adalah salah salah satu metode yang paling populer untuk menentukan jumlah *cluster* yang optimal. Cara untuk menentukan jumlah *cluster* terbaik dengan metode *elbow* yaitu dengan mengamati hasil perbandingan grafik antara setiap jumlah *cluster* k yang akan membentuk perbedaan paling mencolok yaitu membentuk sudut siku atau *elbow* pada suatu titik. Setiap jumlah *cluster* k yang dihasilkan dengan menggunakan metode *K-means* diuji dengan menggunakan persamaan *SSE* (*Sum of Square Error*) yaitu merupakan persamaan yang mengukur perbedaan atau selisih *SSE* tiap jumlah *cluster* k yang diperoleh dengan metode K-means sebelumnya [11].

Berikut ini adalah persamaan yang digunakan untuk menghitung SSE [12]:

$$SSE = \sum_{k=1}^{k} \sum_{x_i} |x_i - c_k|^2 \quad (6)$$

Dimana :
k = cluster ke-c,
x_i = jarak data objek ke-*i*
c_k = pusat cluster ke-*i*

Berikut ini adalah langkah analisis elbow:

1. Melakukan pengelompokan dengan metode *k-means* pada berbagai variasi nilai k yang berbeda. Contohnya, jumlah nilai k dari 1 hingga 7 *cluster*.
2. Pada setiap *cluster* k, tentukan nilai *sum of square error* (*SSE*).
3. Plot kurva *SSE* sesuai dengan jumlah *cluster* k.
4. Titik sudut yang membentuk siku pada plot umumnya ditentukan sebagai indikator jumlah *cluster* terbaik.

2.6. Metode *Silhouette Coefficient*

Silhouette Coefficient adalah salah satu metode yang digunakan untuk mengevaluasi *k-means* atau menentukan jumlah *cluster* yang terbaik. Metode ini adalah kombinasi dari metode *Cohesion* dan *Separation*. Metode *Cohesion* merupakan metode yang digunakan untuk mengukur kedekatan hubungan antara objek dalam satu kelompok yang sama, sedangkan metode *Separation* merupakan metode yang digunakan untuk mengukur jarak antara kelompok terpisah dengan kelompok yang lainnya.

2.7. Segmentasi Pelanggan

Segmentasi pelanggan adalah proses membagi pelanggan menjadi beberapa kelompok berdasarkan kesamaan karakteristik pelanggan seperti demografi atau perilaku pembelian. Proses segmentasi menggambarkan karakteristik kelompok pelanggan (disebut kelompok atau *cluster*) dalam data. Segmentasi berarti menempatkan populasi kedalam kelompok sesuai dengan keterkaitan mereka atau karakteristik mereka yang serupa. [13].

2.8. Analisis RFM

Analisis RFM adalah kependekan dari *recency, frequency, dan monetary*. RFM dapat didefinisikan sebagai alat analisis pemasaran untuk menentukan secara kuantitatif pelanggan mana yang terbaik untuk perusahaan dengan menganalisis dan memeriksa informasi terkait waktu pembelian terakhir (*recency*), berapa kali pelanggan melakukan transaksi (*frequency*), dan berapa total yang dihabiskan pelanggan dalam melakukan pembelian (*monetary*) [14].

Berikut ini adalaha tabel contoh tingkatan *RFM* Pelanggan seperti pada tabel 1.

Tabel 1. Contoh Tingkatan RFM Pelanggan

Recency	*Frequency*	*Monetary*
R-Tier-1 (*most recent*)	F-Tier-1 (*most frequent*)	M-Tier-1 (*highest spend*)
R-Tier-2	F-Tier-2	M-Tier-2
R-Tier-3	F-Tier-3	M-Tier-3
R-Tier-4 (*least recent*)	F-Tier-4 (*only one transaction)*	M-Tier-4 (*lowest spend*)

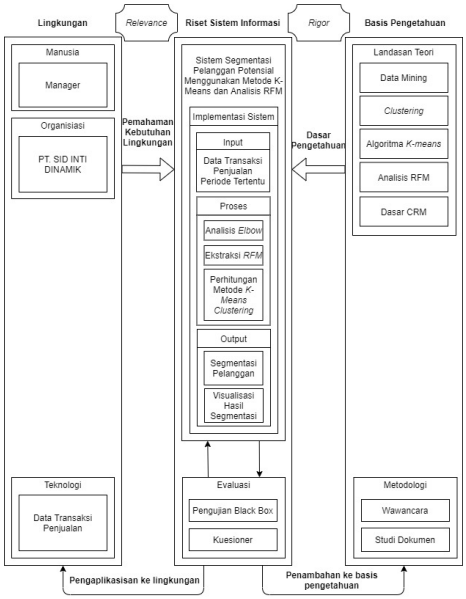
p-ISSN : 2338-493X
e-ISSN : 2809-574X

Berikut ini adalah contoh tipe pelanggan berdasarkan skor rfm:

- *Best Customer* - Kelompok ini terdiri dari pelanggan yang ditemukan di R-Tier-1, F-Tier-1 dan M-Tier-1. Mereka melakukan transaksikan baru-baru ini, sering melakukannya dan membelanjakan lebih banyak daripada pelanggan lain di perusahaan. Kelompok ini akan diberi notasi singkatan 1-1-1; dan akan menggunakan notasi seperti ini kedepannya.
- *High spending New Customers* – Kelompok ini terdiri dari para pelanggan dalam kategori 1-4-1 dan 1-4-2. Ini adalah pelanggan yang telah bertransaksi hanya sekali, tetapi baru-baru ini dan mereka menghabiskan banyak untuk pembelian.
- *Lowest Spending Active Loyal Customers* - Kelompok ini terdiri dari para pelanggan dalam kategori 1-1-3 dan 1-1-4 dan mereka telah bertransaksi baru-baru ini dan sering melakukannya tetapi menghabiskan paling sedikit.
- *Churned Best Customers* - Kelompok ini terdiri dari pelanggan dalam kategori 4-1-1, 4-1-2, 4-2-1 dan 4-2-2 dan mereka telah sering bertransaksi dan menghabiskan banyak, tetapi sudah dalam waktu yang lama sejak transaksi terakhir mereka dengan perusahaan.

3. METODE PENELITIAN

Metodologi penelitian ini menggunakan metodologi dari *Information Systems Research Framework* [15]. *Information Systems Research Framework* adalah *framework* yang berfokus pada *Design Science Research* (*DSR*) yang merupakan panduan penelitian yang membuat dan mengevaluasi suatu penelitian dibidang teknologi informasi untuk memecahkan masalah yang dihadapi organisasi. *DSR* berupaya untuk menciptakan inovasi yang mendefinisikan ide-ide, kemampuan teknis, manajemen, desain, implementasi, serta produk yang dihasilkan melalui analisis agar penggunaan sistem informasi dapat berjalan secara efektif dan efisien. Metodologi dalam penelitian ini dapat dilihat pada gambar 2 di bawah ini:



Gambar 2. Metode Penelitian Berdasarkan *IS Research Framework*

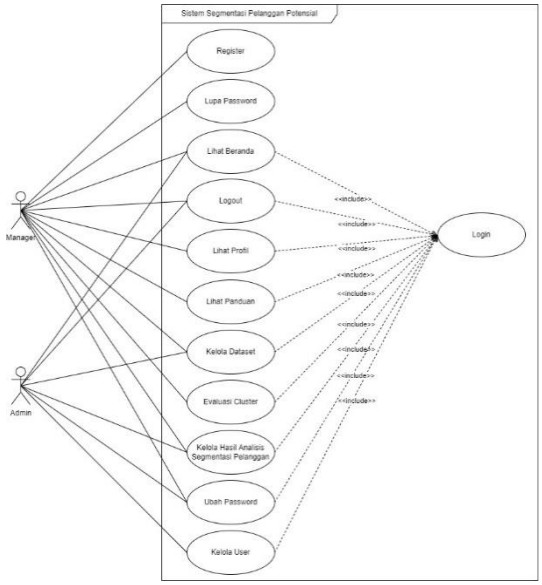
Pada Information Systems Research Framework lingkungan mendefinisikan ruang lingkup masalah yang terdiri orang, organisasi (bisnis), dan teknologi yang ada atau yang direncanakan. Pada penelitian ini orang yang menjadi tujuan adalah manager, lalu organisasi yang menjadi tujuan adalah PT. SID INTI DINAMIK, dan teknologi yang menjadi tujuan adalah data transaksi penjualan.

4. HASIL DAN PEMBAHASAN

4.1. *Use Case Diagram*

Adapun perancangan *Use Case* dari Sistem Segmentasi Pelanggan Potensial dapat dilihat pada gambar 3.

p-ISSN : 2338-493X
e-ISSN : 2809-574X

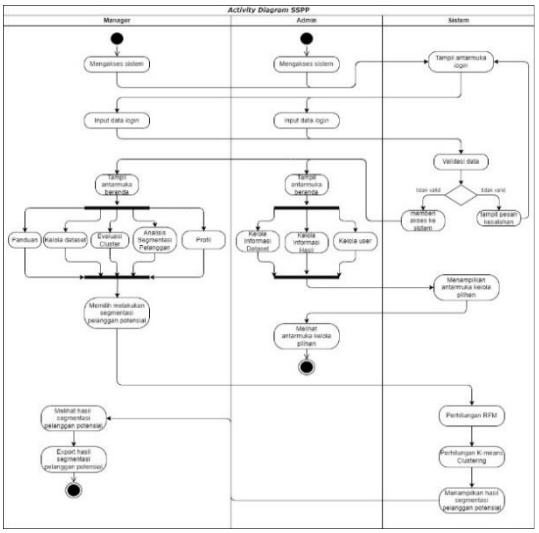


Gambar 3. *Use Case* Diagram

Use Case Diagram adalah sebuah model diagram UML yang secara singkat menggambarkan kebutuhan fungsional, siapa yang akan menggunakan sistem, dan fugsi apa saja yang dapat dilakukan pada sistem yang akan dibagun.

4.2. *Activity Diagram*

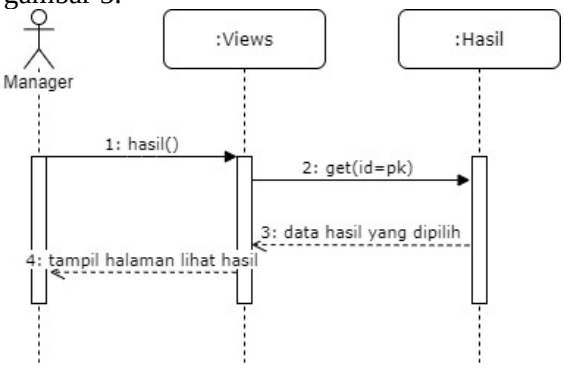
Adapun perancangan Activity Diagram SSPP dalam melakukan perhitungan dapat dilihat pada gambar 4. *Activity Diagram* merupakan sebuah rancangan yang menggambarkan tentang alur kegiatan atau aktivitas dalam sistem yang yang akan dibangun serta bagaimana proses alur dimulai, keputusan yang mungkin terjadi, dan bagaimana sistem akan berakhir.



Gambar 4. *Activity Diagram*

4.3. *Sequence Diagram*

Adapun perancangan *Sequence Diagram* SSPP dalam melakukan tambah hasil analisis segmentasi pelanggan dapat dilihat pada gambar 5.

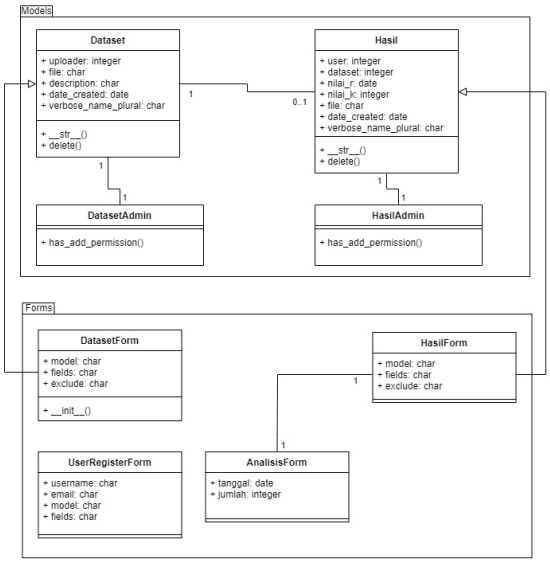


Gambar 5. *Sequence Diagram* Tambah Hasil Analisis Segmentasi Pelanggan

Sequence Diagram merupakan sebuah model diagram UML yang memvisualisasikan interaksi yang terjadi dalam kolaborasi dinamis antara sejumlah objek untuk menunjukkan urutan pesan yang dikirim antar objek.

4.4. *Class Diagram*

Adapun perancangan *Class Diagram* SSPP dapat dilihat pada gambar 6. *Class Diagram* merupakan sebuah model diagram UML yang menggambarkan dengan jelas struktur serta deskripsi kelas, atribut, operasi, hubungan apa yang terjadi dari setiap objek pada sebuah sistem.

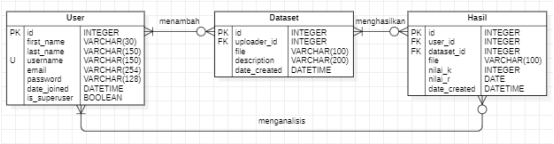


Gambar 6. *Class Diagram*

p-ISSN : 2338-493X
e-ISSN : 2809-574X

4.5. Entity Relationship Diagram

Adapun perancangan *Entity Relationship Diagram* SSPP dapat dilihat pada gambar 7.



Gambar 7. *Entity Relationship Diagram*

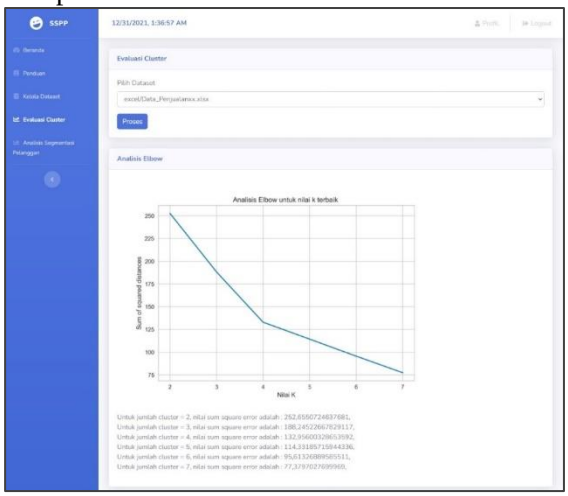
Entity Relationship Diagram atau *ERD* merupakan suatu model diagram struktural yang digunakan dalam desain *database* dimana *ERD* menggambarkan hubungan antar objek-objek data yang memiliki hubungan antar relasi pada sebuah sistem.

4.6. Implementasi

Berikut merupakan implementasi dari fungsional sistem segmentasi pelanggan potensial menggunakan metode *K-means* dan Analisis RFM

a. Halaman Evaluasi Cluster

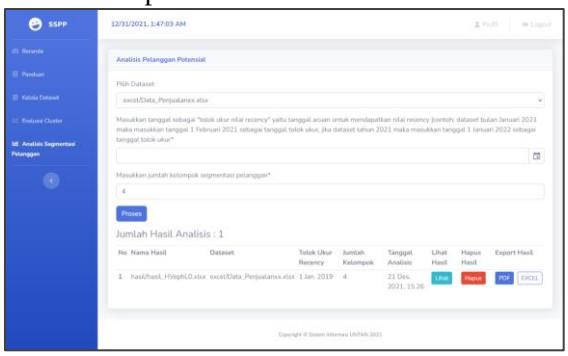
Halaman lihat evaluasi cluster merupakan halaman yang ditemui ketika *user* ingin melakukan evaluasi cluster pada *dataset* yang telah ditambahkan. Gambar 8 adalah tampilan halaman evaluasi cluster setelah diimplementasikan.



Gambar 8. Implementasi Halaman Evaluasi Cluster

b. Halaman Kelola & Tambah Hasil Analisis Segmentasi Pelanggan

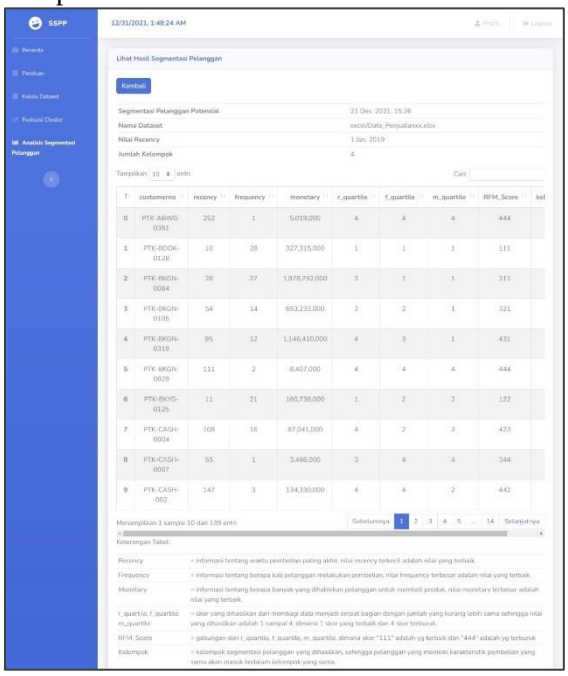
Halaman kelola & tambah hasil analisis segmentasi pelanggan merupakan halaman yang ditemui ketika *user* ingin melakukan analisis segmentasi pelanggan menggunakan *dataset* yang telah ditambahkan sebelumnya. Gambar 9 adalah tampilan halaman kelola & tambah hasil analisis segmentasi pelanggan setelah diimplementasikan.



Gambar 9. Implementasi Halaman Kelola & Tambah Hasil Analisis Segmentasi Pelanggan

c. Halaman Lihat Hasil Analisis Segmentasi Pelanggan

Halaman lihat hasil analisis segmentasi pelanggan merupakan halaman yang ditemui ketika *user* ingin melihat hasil analisis yang telah ditambahkan sebelumnya. Gambar 10 adalah tampilan halaman lihat hasil analisis segmentasi pelanggan setelah diimplementasikan.



Gambar 10. Implementasi Halaman Lihat Hasil Analisis Segmentasi Pelanggan

5.1. Pembahasan

Berikut merupakan implementasi dari setiap tahapan *data mining* berdasarkan tahapan

p-ISSN : 2338-493X
e-ISSN : 2809-574X

KKD (*Knowledge Discovery in Database*) dari sistem segmentasi pelanggan potensial menggunakan metode K-means dan Analisis RFM.

a. *Data Selection*

Tabel 2. Sampel Dataset

no	Customerno	customername	product	invoiceno	invoicedate	description	amount
1	PTK-BKGN-0084	Karya Tani / Ngateman	PESTISIDA	5533	1/10/2018		176819000
2	PTK-BKGN-0084	Karya Tani / Ngateman	PESTISIDA	5533	1/10/2018		176819000
3	PTK-JNGT-0057	Duta Tani	PESTISIDA	5476	1/10/2018		4114000
4	PTK-JNGT-0057	Duta Tani	SPRAYER	5477	1/10/2018		473000
5	PTK-JNGT-0067	Hira Agrotama	PESTISIDA	5475	1/10/2018		2182000
6	PTK-JNGT-0108	Mustajab	PESTISIDA	5478	1/10/2018		1564000
7	PTK-KKAP-0060	H.Madsah	PESTISIDA	5412	1/10/2018		10071000
8	PTK-KKAP-0096	Mawar Tani	PESTISIDA	5390	1/10/2018		56000000
9	PTK-KKAP-0118	A. Santoso	PESTISIDA	5429	1/10/2018		3452000
10	PTK-KKAP-0118	A. Santoso	PESTISIDA	5434	1/10/2018		2119000
..	…	…	…	…	…		…
2194	PTK-KTPG-0130TN	CV. SAHABAT TANI JAYA	PESTISIDA	11755	25/12/2018		31419000

Tahap pertama dalam implementasi *data mining* adalah tahap *data selection* atau pemilihan data. Data yang dipilih haruslah memenuhi kebutuhan dari proses *data mining* yang akan dilakukan, dimana atribut yang butuhkan adalah customerno, customername, invoiceno, invoicedate, dan amount sehingga data yang dipilih haruslah memiliki atribut tersebut. Data yang dipilih dapat dalam rentang waktu kapanpun seperti 1 bulan, 3 bulan, atau 1 tahun sesuai kebutuhan pengguna.

Pada penelitian ini data yang dipilih adalah data transaksi penjualan periode 2018 yang digunakan sebagai *dataset* dalam implementasi *data mining* dimana dataset tersebut memiliki 2194 baris data. Berikut adalah 10 baris data sampel dari dataset yang digunakan dapat dilihat pada tabel 2.

Tabel 2 merupakan sampel dari dataset yang akan diupload kedalam sistem melalui halaman kelola dataset.

b. *Pre-processing /cleaning*

Setelah dataset dipilih, tahap selanjutnya adalah *pre-processing* data atau *cleaning*. Proses *pre-processing* dilakukan oleh sistem untuk pengecekan pada dataset apakah dataset yang diupload memiliki kolom yang dibutuhkan dalam proses *data mining* yaitu kolom customerno, customername, invoiceno, invoicedate, dan amount, jika tidak memiliki kolom tersebut maka proses *data mining* tidak dapat dilanjutkan karena tidak sesuai dengan ketentuan sistem. Jika dataset tersebut memenuhi syarat maka proses pre-processing akan dilanjutkan ke proses pembersihan data yang dilakukan untuk dan menghilangkan data yang duplikat dan membersihkan baris data yang tidak efisien seperti baris data yang tidak memiliki nilai.

c. *Transformation*

Transformasi yang dilakukan pada tahap ini yaitu analisis RFM dan mencari skor RFM setiap pelanggan. Dataset akan otomatis ditransformasi oleh sistem saat proses analisis segmentasi pelanggan potensial dilakukan.

Analisis RFM adalah proses yang dilakukan untuk menentukan nilai atribut RFM yaitu *recency*, *frequency* dan *monetary*. Nilai *recency* merupakan selisih waktu yang ditentukan sebagai tolok ukur yaitu 1 Januari 2019 dengan waktu terakhir melakukan transaksi. Tanggal 1 Januari 2019 dipilih karena

p-ISSN : 2338-493X
e-ISSN : 2809-574X

dataset yang digunakan merupakan data penjualan tahun 2018 sehingga digunakan tanggal tersebut untuk mengetahui segmentasi pelanggan potensial pada tahun 2018. Nilai *recency* didapatkan dari tanggal tolok ukur perhitungan dikurang tanggal yang paling akhir dari transaksi setiap pelanggan.

Setelah mendapatkan nilai *recency*, *frequency* dan *monetary* pada masing-masing pelanggan, transformasi tahap selanjutnya adalah mencari skor RFM dengan menggunakan kuartil /*quartile* sehingga didapatkan 4 kelompok tingkatan nilai pada setiap rfm dimana skor 1 adalah skor terbaik dan skor 4 adalah terburuk. Kuartil membagi sekelompok data yang telah diurutkan menjadi empat bagian yang sama besar. Ada tiga buah kuartil yang diperoleh dari suatu himpunan data yaitu kuartil 1 (Q1), kuartil 2 (Q2) atau median, dan kuartil 3 (Q3).

d. *Data Mining*

Setelah melalui tahap *transformation* atau transformasi data, tahap selanjutnya adalah tahap *data mining*. Pada tahap ini proses *data mining* yang dilakukan adalah proses *clustering* menggunakan algoritma *K-means*. Proses ini dimulai dengan menentukan jumlah kelompok yang diinginkan. Data yang akan digunakan dalam perhitungan ini adalah skor RFM setiap pelanggan hasil dari transfromasi data.

Berikut ini adalah sampel baris data hasil dari simulasi perhitungan dengan menggunakan metode K-means clustering pada Tabel 3.

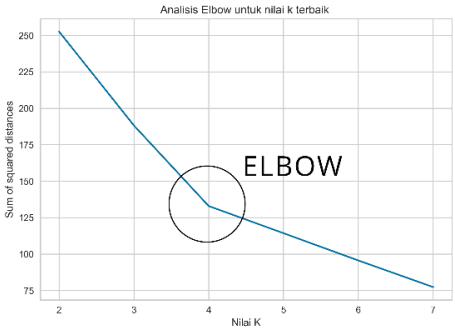
Tabel 3. Sampel Hasil Perhitungan

Data ke.	customerno	R Score	F Score	M Score	Cluster
1	PTK-ABWG-0351	4	4	4	3
2	PTK-BDOK-0128	1	1	1	1
3	PTK-BKGN-0084	3	1	1	1
4	PTK-BKGN-0105	3	2	1	2
5	PTK-BKGN-0319	4	3	1	2
6	PTK-BKGN-0829	4	4	4	3
7	PTK-BKYG-0125	1	2	2	1
8	PTK-CASH-0004	4	2	3	2
9	PTK-CASH-0007	3	4	4	3
10	PTK-CASH-002	4	4	2	2
…	…	…	…	…	…
139	PTK-TYAN-0234	2	2	3	4

e. *Interpretation /evaluation*

Interpretation atau *evaluation* adalah tahap menampilkan pola atau informasi yang didapatkan dari proses *data mining* ke dalam bentuk yang lebih mudah dimenherti oleh pihak yang berkepentingan. Pada tahap ini akan dilakukan visualisasi pada data hasil dari segmentasi pelanggan potensial.

Berikut ini adalah grafik elbow yang dihasilkan dari perhitungan menggunakan sistem segmentasi pelanggan potensial.

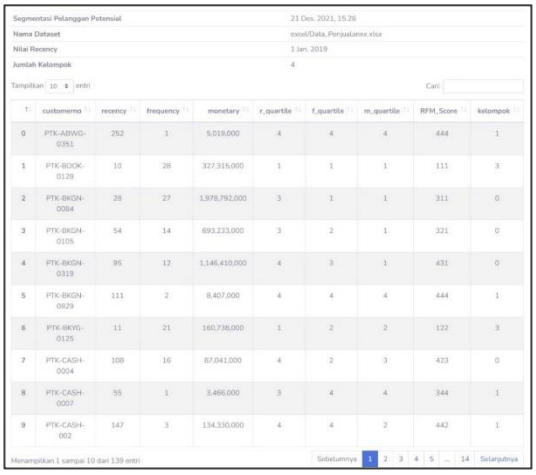


Gambar 11. Grafik elbow hasil perhitungan sistem

Pada gambar 11 grafik tersebut menampilkan hasil setiap perhitungan berdasarkan jumlah kelompok antara 2 hingga 7, nilai *cluster* k 3 ke nilai *cluster* k 4 terlihat penurunan drastis membentuk siku (*elbow*) pada titik k 4 ke k 5 dan penurunan seterusnya ke nilai berikutnya sudah cukup kecil. Berdasarkan pada grafik tersebut terlihat garis yang membentuk siku pada sudut yang bertitik pada kelompok 4 maka berdasarkan hasil pengamatan tersebut maka jumlah kelompok yang ideal adalah k=4, sehingga hasil clustering akan menghasilkan 4 kelompok.

Berikut ini adalah gambar dari hasil segmentasi pelanggan menggunakan sistem segmentasi pelanggan potensial yang menghasilkan 4 cluster atau kelompok:

p-ISSN : 2338-493X
e-ISSN : 2809-574X



Segmentasi Pelanggan Potensial	21 Des. 2021, 15.26
Nama Dataset	excel/Data_Penjualanxx.xlsx
Nilai Recency	1 Jan. 2019
Jumlah Kelompok	4

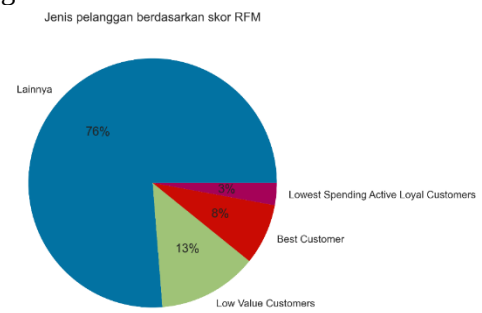
Tampilkan 10 entri Cari:

↑↓	customerno	recency	frequency	monetary	r_quartile	f_quartile	m_quartile	RFM_Score	kelompok
0	PTK-ABWG-0351	252	1	5,019,000	4	4	4	444	1
1	PTK-BOOK-0128	10	28	327,315,000	1	1	1	111	3
2	PTK-BKGN-0084	28	27	1,978,792,000	3	1	1	311	0
3	PTK-BKGN-0106	54	14	693,233,000	3	2	1	321	0
4	PTK-BKGN-0319	95	12	1,146,410,000	4	3	1	431	0
5	PTK-BKGN-0829	111	2	8,407,000	4	4	4	444	1
6	PTK-BKYG-0125	11	21	160,738,000	1	2	2	122	3
7	PTK-CASH-0004	108	16	87,041,000	4	2	3	423	0
8	PTK-CASH-0007	55	1	3,466,000	3	4	4	344	1
9	PTK-CASH-002	147	3	134,330,000	4	4	2	442	1

Menampilkan 1 sampai 10 dari 139 entri Sebelumnya 1 2 3 4 5 ... 14 Selanjutnya

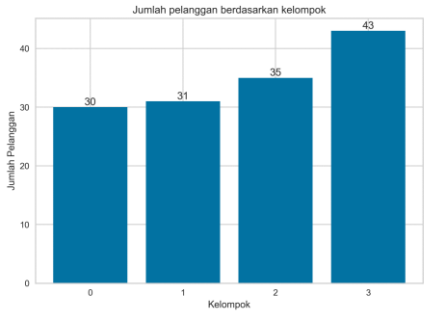
Gambar 12. Hasil segementasi pelanggan menggunakan sistem

Gambar 12 merupakan hasil segmentasi pelanggan menggunakan sistem segmentasi pelanggan potensial. Berdasarkan hasil tersebut, didapatkan pelanggan dengan skor rfm 111 menepati posisi kelompok atau cluster ke 3. Berdasarkan hasil dari segmentasi pelanggan yang dihasilkan maka dapat dilakukan visualisasi data agar informasi yang dihasilkan dari proses *data mining* dapat lebih mudah dimengerti. Berikut ini adalah visualisasi data yang dihasilkan:



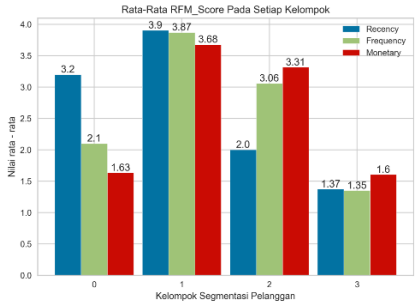
Gambar 13. Visualisasi 1

Gambar 13 merupakan *pie chart* yang memberi informasi tentang persentase tipe pelanggan berdasarkan skor rfm. Berdasarkan hasil dari skor rfm menghasilkan 13% pelanggan termasuk kedalam tipe *low value customers*, 8% pelanggan termasuk kedalam tipe *best customer*, 3% pelanggan termasuk kadalam *lowest spending active loyal customers*, dan 78% pelanggan termasuk kedalam kelompok lainnya.



Gambar 14. Visualisasi 2

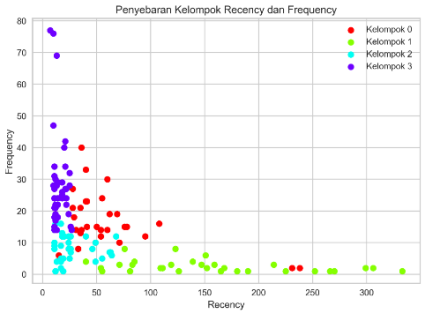
Gambar 14 merupakan *bar chart* yang memberi informasi tentang banyaknya anggota setiap kelompok atau *cluster* yang dihasilkan. Berdasarkan hasil dari segmentasi pelanggan potensial yang menghasilkan 4 kelompok yaitu kelompok 0, 1, 2 dan 3, kelompok 0 memiliki sebanyak 30 anggota, kelompok 1 memiliki sebanyak 31 anggota, kelompok 2 memiliki sebanyak 35 anggota, dan kelompok 3 memiliki sebanyak 43 anggota. Kelompok yang memiliki anggota terbanyak adalah kelompok 3, dan anggota tersedikit adalah kelompok 0.



Gambar 15. Visualisai 3

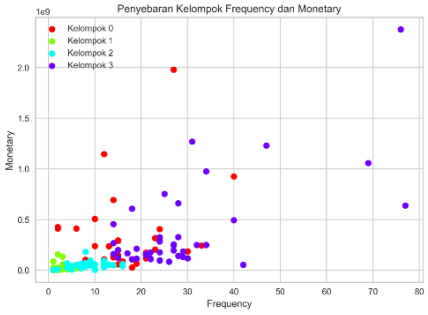
Gambar 15 merupakan *bar chart* yang memberi informasi tentang nilai rata-rata atau *mean* dari skor RFM setiap kelompok. Berdasarkan hasil dari segmentasi pelanggan potensial yang menghasilkan 4 kelompok yaitu kelompok 0, 1, 2 dan 3, kelompok 3 memiliki nilai rata-rata skor RFM yang terendah dan kelompok 1 memiliki nilai rata-rata skor RFM tertinggi. Maka kelompok pelanggan yang terbaik adalah kelompok dengan nilai rata-rata terendah dan kelompok pelanggan terburuk adalah kelompok dengan nilai rata-rata tertinggi.

p-ISSN : 2338-493X
e-ISSN : 2809-574X



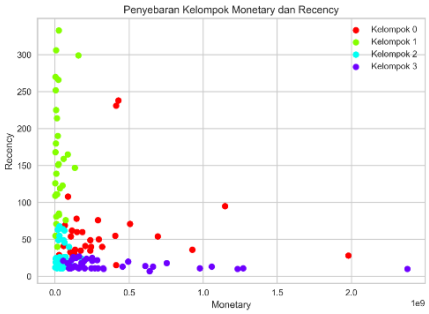
Gambar 16. Visualisasi 4

Gambar 16 merupakan *scatter plot* yang memberi informasi tentang perbandingan penyebaran kelompok berdasarkan nilai *frequency* dan *recency*. Berdasarkan visualisasi tersebut dapat diamati pada kelompok 3 sebagai kelompok terbaik cenderung mempunyai nilai *frequency* yang besar dan nilai *recency* yang kecil. Maka kelompok pelanggan yang terbaik adalah kelompok 3, karena kelompok 3 mempunyai nilai *frequency* yang besar dan nilai *recency* yang kecil pada anggota kelompoknya dibandingkan pada ketiga kelompok lainnya. dikarenakan semakin kecil nilai *recency* maka semakin baik.



Gambar 17. Visualisasi 5

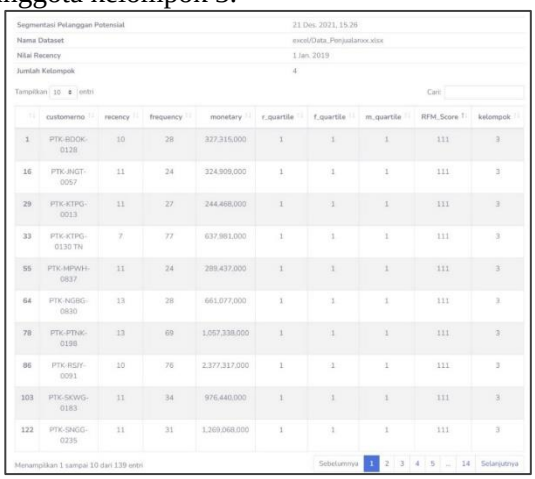
Gambar 17 merupakan *scatter plot* yang memberi informasi tentang perbandingan penyebaran kelompok berdasarkan nilai *monetary* dan *frequency*. Berdasarkan visualisasi tersebut dapat diamati pada kelompok 3 sebagai kelompok terbaik berdasarkan visualisasi sebelumnya cenderung mempunyai nilai *monetary* dan *frequency* yang besar. Maka kelompok pelanggan yang terbaik adalah kelompok 3, karena kelompok 3 memiliki nilai *monetary* dan *frequency* yang lebih besar pada anggota kelompoknya dibandingkan pada ketiga kelompok lainnya.



Gambar 18. Visualisasi 6

Gambar 18 merupakan *scatter plot* yang memberi informasi tentang perbandingan penyebaran kelompok berdasarkan nilai *recency* dan *monetary*. Berdasarkan visualisasi tersebut dapat diamati pada kelompok 3 sebagai kelompok terbaik berdasarkan visualisasi sebelumnya cenderung mempunyai nilai *recency* yang kecil dan nilai *monetary* yang besar. Maka kelompok pelanggan yang terbaik adalah kelompok 3, karena kelompok 3 mempunyai nilai *recency* yang kecil dan nilai *monetary* yang lebih besar pada anggota kelompoknya, dimana semakin kecil nilai recency maka semakin baik.

Berdasarkan pengamatan dari hasil segmentasi pelanggan potensial dapat diambil kesimpulan bahwa kelompok 3 yang memiliki 43 anggota merupakan kelompok pelanggan potensial, berikut adalah sampel data dari anggota kelompok 3.



Segmentasi Pelanggan Potensial — 21 Des. 2021, 15.26
Nama Dataset — excel/Data_Penjualanxx.xlsx
Nilai Recency — 1 Jan. 2019
Jumlah Kelompok — 4

Tampilkan 10 entri — Cari:

	customerno	recency	frequency	monetary	r_quartile	f_quartile	m_quartile	RFM_Score	kelompok
1	PTK-BOOK-0128	10	28	327,315,000	1	1	1	111	3
16	PTK-JNGT-0057	11	24	324,909,000	1	1	1	111	3
29	PTK-KTPG-0013	11	27	244,468,000	1	1	1	111	3
33	PTK-KTPG-0130 TN	7	77	637,981,000	1	1	1	111	3
55	PTK-MPWH-0837	11	24	289,437,000	1	1	1	111	3
64	PTK-NGBG-0830	13	28	661,077,000	1	1	1	111	3
78	PTK-PTNK-0198	13	69	1,057,338,000	1	1	1	111	3
86	PTK-RSJY-0091	10	76	2,377,317,000	1	1	1	111	3
103	PTK-SKWG-0183	11	34	976,440,000	1	1	1	111	3
122	PTK-SNGG-0235	11	31	1,269,068,000	1	1	1	111	3

Menampilkan 1 sampai 10 dari 139 entri — Sebelumnya 1 2 3 4 5 ... 14 Selanjutnya

Gambar 19. Hasil segmentasi pelanggan kelompok 3

5. KESIMPULAN

a. Dengan adanya sistem segmentasi pelanggan potensial berbasis web yang menerapkan metode clustering algoritma

p-ISSN : 2338-493X
e-ISSN : 2809-574X

K-Means dan analisis RFM, dapat menjadi alat bantu untuk mendapatkan rekomendasi dalam analisis dan mencari segmentasi pelanggan potensial berdasarkan data penjualan. Berdasarkan hasil pengujian fungsional sistem telah dilakukan oleh pihak PT. SID INTI DINAMIK dan pihak lainnya sebagai sasaran pengguna sistem, sistem berbasis web yang dibangun dapat berjalan sesuai dengan kebutuhan yang telah ditentukan. Dan berdasarkan hasil pengujian antarmuka sistem yang telah dilaksanakan kepada masyarakat umum, sistem dikategorikan sangat baik dengan persentase nilai akhir yaitu 87,59%.

b. Dengan menerapkan metode *K-means clustering* dapat digunakan untuk mengelompokkan pelanggan ke dalam kelompok yang memiliki karakteristik serupa. Analisis RFM dapat menjadi variabel yang digunakan dalam melakukan segmentasi pelanggan potensial menggunakan metode *K-means clustering*. Berdasarkan simulasi perhitungan yang dilakuakn, didapatkan hasil sebanyak 4 kelompok yaitu kelompok 0, 1, 2 dan 3. Kelompok pelanggan potensial terbaik adalah kelompok 3 dengan jumlah anggota sebanyak 43 dengan skor RFM "111" dimana skor rata-rata kelompok yaitu skor recency 1.47, skor frequency 1.35, dan skor monetary 1.6.

c. Dengan adanya visualisasi hasil segmentasi pelanggan berbasis web yang ditampilkan menggabungkan beberapa grafik sehingga lebih interaktif, dapat membantu penyajian data dan informasi sehingga lebih mudah dipahami dalam membaca hasil segmentasi pelanggan.

7. SARAN

Adapun saran yang akan diberikan adalah sebagai berikut:

a. Penelitian selanjutnya diharapkan dapat menambahkan variabel lainnya selain variabel RFM untuk pengembangan penelitian yang lebih lanjut. Contoh variabel yang dapat digunakan yaitu jarak waktu dari transaksi pertama hingga transaksi terakhir.

b. Penelitian selanjutnya diharapkan dapat melakukan evaluasi tingkat akurasi dari hasil *clustering* yang telah dilakukan dengan menggunakan metode pengujian tingkat akurasi untuk metode k-means *clustering.*

c. Berdasarkan hasil pengujian pada sistem segmentasi pelanggan potensial (SSPP) ini terdapat kekurangan pada panduan penggunaan yang masih sulit untuk dimengerti oleh pengguna awam. Penelitian selanjutnya diharapkan dapat mengembangkan sistem segmentasi pelanggan potensial dengan memberikan panduan penggunaan sistem yang lebih informatif. Contohnya seperti menggunakan bahasa yang lebih mudah dimengerti dan menggunakan video atau infografik panduan penggunaan sistem.

DAFTAR PUSTAKA

[1] S. Wijaya, and S. Thio, "Implementasi Membership Card dan Pengaruhnya Dalam Meningkatkan Loyalitas Pengunjung Restoran di Surabaya," *Jurnal Eksekutif*, vol. 3, no. 2, 2007.

[2] N. Herawati and M. Sulistyowati, "Strategi Meningkatkan Kepuasan dan Loyalitas Pelanggan Melalui Citra Merek, Kualitas Produk dan Promosi Pada Pelanggan Smarthphone Samsung di Surakarta," *Probank: Jurnal Ekonomi dan Perbankan*, Vol 5, No 2, pp. 214-226, 2020.

[3] Magento. *Identifying Your Most Profitable Customers: An Introduction to Customer Segmentation.* (2019). Accessed: May. 28, 2021. [Online]. Available: https://magento.com/sites/default/files8/2019-01/introduction-to-customer-segmentation-v2.pdf

[4] E. Turban, J. E. Aronson, and T. P. Liang, *Decision Support Systems and Intelligent Systems*. U.S.A: Prentice-Hall, Inc, 2005.

[5] K. Fatmawati and A. P. Windarto, "Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (DBD) Berdasarkan Provinsi," *CESS (Journal of Computer Engineering System and Science)*, vol. 3, no. 2, pp. 173-178, 2018.

[6] D. D. C. Nugraha, Z. Naimah, M. Fahmi, Makhfuzi, and N. Setiani, "Klasterisasi

p-ISSN : 2338-493X
e-ISSN : 2809-574X

Judul Buku dengan Menggunakan Metode K- Means," *in Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, 21 Juni 2014, pp. G-1 – G-4.

[7] P. N. Tan, M. Steinbach, and V. Kumar, *Introduction Data Mining*. Edinburgh Gate, England: Pearson Education Limited, 2014.

[8] D. T. Larose and C. D. Larose. *Data Mining and Predictive Analytics*. Hoboken, New Jersey: John Wiley & Sons, Inc, 2015.

[9] A. N. Khomarudin. *Teknik Data Mining : Algoritma K-Means Clustering*. Komunitas eLearning Ilmu Komputer. 2018.

[10] A. S. Pratikno, A. A. Prastiwi, and S. Rahmawati, "Kuartil, Desil, dan Presentil Serta Cara Menghitungnya dalam Distribusi Frekuensi" *OSF Preprint*, vol. 3, 2020.

[11] N. P. E. Merliana, Ernawati., and A. J. Santoso, "Analisa Penentuan Jumlah Cluster Terbaik Pada Metode K-Means Clustering," *Proceeding SENDI_U Ke-2*, 2015.

[12] D. A. I. Cahya and D. A. K. Pramita, "Analisis Perbandingan Metode Elbow dan Sillhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," *Matrix: Jurnal Manajemen Teknologi Dan Informatika*, vol. 9, no. 3, pp. 102-109, 2019.

[13] S.M.H. Jansen. *Customer Segmentation and Customer Profiling for a Mobile Telecommunications Company Based on Usage Behavior*. Accessed : May 16, 2021 [online]. Available : https://www.semanticscholar.org/paper/Customer-Segmentation-and-Customer-Profiling-for-a -Jansen/7a3a688783e0424bd89f7413138bbfc24deeef8f

[14] H. Bhasin, *What is RFM Analysis (Recency, Frequency & Monetary Value)?* Accessed : May 28, 2021, Available : https://www.marketing91.com/what-is-rfm-model.

[15] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design Science in Information Systems Research," *MIS Quarterly*, vol. 28, no. 1, pp. 75–105, 2004.