

# Analisis *Clustering* Pelanggan Berdasarkan Data Transaksi Penjualan Menggunakan Algoritme *K-Means* dan Metode *Recency, Frequency, Monetary* (RFM) (Studi Kasus: CV XYZ)

Yosi Yonata<sup>#1</sup>, Herastia Maharani<sup>#2</sup>, Chen Viona<sup>#3</sup>

<sup>#</sup>Program Studi Sistem Informasi, Institut Teknologi Harapan Bangsa  
Jalan Dipatiukur No. 84-86, Bandung, Indonesia 40132

<sup>1</sup>yosi@ithb.ac.id

<sup>2</sup>herastia@ithb.ac.id

<sup>3</sup>vionnachen@gmail.com

**Abstract**— CV XYZ is a company engaged in the retail of Muslim women's clothing and accessories, located in Bandung. The products sold consist of 5 types of categories. Sales business processes that occur in the company generate sales transaction data. CV XYZ expects the transaction data it has can be processed and used to get information about customer segmentation using the suitable method. This research utilizes the RFM method at the pre-processing stage and uses the clustering method at the next stage. Determination of the algorithm and the number of clusters obtained from the evaluation process using the standard deviation technique and the Davies Bouldin Index method. As a result, *K-means* is the most appropriate algorithm to be applied to CV XYZ transaction data by dividing customers into 3 clusters.

**Keywords**— retail, transaction data, purchase, RFM, clustering, *K-means*, potential customer

**Abstrak**— CV XYZ adalah sebuah perusahaan yang bergerak di bidang retail pakaian dan aksesoris busana muslim wanita yang berlokasi di Bandung. Produk yang dijual terdiri dari 5 jenis kategori. Proses bisnis penjualan yang terjadi di perusahaan menghasilkan data transaksi penjualan. CV XYZ mengharapkan data transaksi yang dimiliki dapat diolah dan dimanfaatkan untuk mendapatkan informasi mengenai segmentasi pelanggan dengan menggunakan metode yang tepat. Penelitian ini memanfaatkan Metode RFM pada tahap *pre-processing* dan menggunakan metode *clustering* pada tahap berikutnya. Penentuan algoritme dan jumlah *cluster* diperoleh dari proses evaluasi yang menggunakan teknik simpangan baku dan metode Davies Bouldin Index. Hasilnya, *K-means* adalah algoritme yang paling tepat untuk diterapkan pada data transaksi CV XYZ dengan membagi pelanggan menjadi 3 *cluster*.

**Kata Kunci**— retail, data transaksi, penjualan, RFM, clustering, *K-means*, pelanggan potensial

## I. PENDAHULUAN

Perusahaan retail merupakan salah satu perusahaan yang melakukan proses bisnis penjualan dengan melibatkan banyak produk dan banyak kategori pelanggan, baik itu dari kategori usia, gender, kota tempat pelanggan tersebut tinggal, dan lain

sebagainya. Proses bisnis tersebut akan menghasilkan data-data seperti data transaksi penjualan, data pelanggan, dan data lainnya. Setiap data tersebut akan disimpan oleh perusahaan dan digunakan untuk mendapatkan informasi yang dibutuhkan. Dalam perusahaan retail, pengumpulan dan pencatatan data merupakan faktor yang paling penting untuk mengembangkan bisnis ke arah yang lebih baik dan membuat bisnis menjadi lebih sukses [1].

Pada penelitian ini studi kasus yang digunakan adalah CV XYZ. CV XYZ merupakan salah satu perusahaan di bidang retail yang memiliki proses penjualan dengan melibatkan banyak jenis produk dan pelanggan. Produk-produk yang dimiliki dan dijual oleh perusahaan adalah produk busana muslim dan juga aksesoris busana muslim. Produk yang dijual dengan merk XYZ ini diminati oleh berbagai pelanggan di berbagai kota dan kabupaten di seluruh Indonesia. Proses jual-beli akan dilakukan secara langsung atau melalui media sosial, seperti Whatsapp. Perusahaan akan mencatat setiap transaksi penjualan yang terjadi secara manual pada bon. Data transaksi yang disimpan dan dimiliki akan menjadi bahan uji pada penelitian ini.

Pencatatan transaksi penjualan di CV XYZ memiliki beberapa data seperti data produk, jumlah produk yang dibeli oleh pelanggan, dan beberapa identitas dari pelanggan tersebut. Pencatatan transaksi ini digunakan sebagai *invoice* atau bukti penjualan barang kepada pelanggan. Setiap salinan bukti tersebut disimpan berdasarkan bulan terjadinya transaksi.

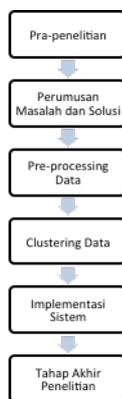
Data transaksi penjualan yang dimiliki dan disimpan oleh perusahaan retail dapat dimanfaatkan untuk beberapa hal, seperti meningkatkan afinitas *brand* perusahaan, mengembangkan dan meningkatkan loyalitas pelanggan, memberikan dan meningkatkan layanan yang lebih baik kepada pelanggan, membantu perusahaan dalam meluncurkan produk baru sesuai dengan kebutuhan dan keinginan pasar, serta memasarkan produk atau *brand* kepada pelanggan ke tingkat selanjutnya [2]. Selain itu, data transaksi penjualan dapat digunakan sebagai sarana berkomunikasi dengan pelanggan agar meningkatkan rasa loyal pada perusahaan. Data juga digunakan untuk menentukan rekomendasi produk dan promosi penjualan pada masing-masing pelanggan [3].

Berdasarkan beberapa manfaat yang dapat dihasilkan dari data transaksi penjualan, data transaksi CV XYZ akan digunakan dan diolah untuk mencapai manfaat tersebut. Salah satu pengolahannya adalah menggunakan metode *data mining* untuk melakukan pengelompokan berdasarkan data pelanggan yang ada atau disebut juga dengan segmentasi pelanggan. Segmentasi pelanggan akan didapatkan berdasarkan karakteristik hasil pengolahan atribut-atribut yang ada pada data transaksi penjualan. Segmentasi pelanggan memberikan peluang kepada CV XYZ untuk menerapkan salah satu konsep *customer relationship management*, di mana perusahaan dapat memberikan pelayanan yang berbeda untuk setiap pelanggan sesuai dengan kebutuhannya [4]. Hal tersebut dapat menjadi peluang perusahaan untuk meningkatkan pelayanan kepada pelanggan sesuai dengan misi CV XYZ.

## II. METODOLOGI

Rumusan masalah dalam penelitian ini adalah bagaimana memanfaatkan data transaksi untuk mendapatkan pengetahuan mengenai segmentasi pelanggan di CV XYZ. Tujuan penelitian ini adalah melakukan segmentasi pelanggan di CV XYZ untuk memanfaatkan data transaksi perusahaan. Penelitian ini akan dilakukan dengan tahapan seperti pada Gambar 1.

- 1) *Tahap Pra-Penelitian* dilakukan dengan melakukan studi literatur, wawancara, dan observasi untuk mendapatkan informasi mengenai proses pengolahan data yang telah berjalan dan pencatatan data yang dilakukan perusahaan. Wawancara dan observasi dilakukan untuk memperoleh hal-hal yang dibutuhkan atau berguna dalam penerapan metode data mining.
- 2) *Perumusan masalah dan solusi* dilakukan untuk menentukan ruang lingkup masalah yang akan diselesaikan, sesuai dengan hasil temuan dari wawancara yang telah dilakukan. Selain itu, dilakukan pula analisis metode untuk menentukan metode yang digunakan dalam pengolahan data perusahaan sehingga dapat mencapai tujuan yang diinginkan.



Gambar 1 Tahapan Penelitian

- 3) *Pre-processing data* dilakukan untuk menentukan yang akan digunakan dan yang bermanfaat dari data yang dimiliki oleh perusahaan.
- 4) *Clustering data* dilakukan dengan menentukan algoritme dan jumlah segmentasi yang tepat sesuai dengan data yang digunakan dan tujuan yang ingin dicapai.
- 5) *Implementasi sistem* dilakukan untuk menghasilkan rancangan sistem yang sesuai dengan tujuan yang telah ditentukan. Selanjutnya akan dilakukan pengujian sistem untuk memastikan bahwa fungsi yang terdapat pada sistem yang dirancang dapat berjalan sesuai dengan kebutuhan dan dapat menghasilkan hasil yang tepat.
- 6) *Tahap akhir penelitian* menghasilkan kesimpulan dari penelitian yang telah dilakukan berdasarkan tujuan dan analisis.

### A. Data Mining

*Data mining* adalah studi untuk mengumpulkan, membersihkan, mengolah, menganalisis, dan memperoleh manfaat informasi dari himpunan data. Banyaknya data yang dihasilkan pada zaman ini membuat seseorang mengolah dan menggunakannya untuk tujuan tertentu [5]. Pada saat penggunaannya, *data mining* memiliki beberapa tahapan yang harus dilakukan. Tahapan tersebut terdiri dari *data cleaning, integration, selection, dan transformation*. Tahapan dilanjutkan dengan proses *data mining*. Terakhir adalah *evaluation dan presentation*. Penerapan *data mining* dapat dilakukan dengan berbagai metode yang tersedia, yaitu *classification, clustering, dan association rules*. Penelitian ini akan menggunakan metode *clustering*.

*Clustering* adalah metode pengelompokan data ke dalam kelompok yang memiliki tingkat kesamaan atau kemiripan. Data yang berbeda akan masuk ke dalam kelompok yang berbeda. Metode ini diterapkan dalam *health psychology, market research, image segmentation, dan lain sebagainya*. Pada penelitian ini, *clustering* yang dilakukan akan menggunakan algoritme *K-means*.

*K-means* adalah algoritme pengelompokan yang paling banyak digunakan untuk mengelompokkan data. *K-means* dimulai dengan memilih titik representative, atau *centroid*, yang akan digunakan sebagai pusat awal dalam pengelompokan data. Rumus untuk menghitung *centroid* menggunakan algoritme *K-means* adalah sebagai berikut [6]:

$$d_{(p,q)} = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} \quad (1)$$

### B. Recency, Frequency, Monetary (RFM)

*Recency, Frequency, Monetary* (RFM) merupakan suatu metode yang digunakan oleh perusahaan ritel untuk memahami perilaku pelanggan. Metode ini melibatkan perhitungan *recency, frequency, dan monetary*. Perhitungan *recency* akan menunjukkan waktu terakhir transaksi pembelian pelanggan di perusahaan tersebut. Perhitungan *frequency* akan menunjukkan jumlah transaksi pembelian pelanggan,

sedangkan perhitungan *monetary* akan menunjukkan nilai pembelian setiap pelanggan.

1. *Recency*

*Recency* pembelian diperoleh dengan menghitung tanggal akhir keseluruhan periode dikurangi tanggal akhir transaksi setiap pelanggan.

2. *Frequency*

*Frekuensi* pembelian diperoleh dengan menghitung rata-rata transaksi yang dilakukan masing-masing pelanggan setiap bulannya.

3. *Monetary*

*Monetary* dilakukan untuk dapat mengetahui seberapa banyak seorang pelanggan menghabiskan uang yang dimiliki untuk membeli dan melakukan transaksi di perusahaan. *Monetary* diperoleh dengan menghitung rata-rata nominal pembelian pelanggan setiap bulannya.

III. HASIL DAN PEMBAHASAN

A. *Pre-processing Data*

Proses pengolahan data dimulai dengan *pre-processing data* yang terbagi empat tahap, yaitu *data cleaning, data integration, data reduction, dan data transformation*. Penelitian ini menggunakan data transaksi penjualan milik CV XYZ dari bulan Juli 2019 hingga Februari 2020. Proses *data cleaning* sendiri terbagi dari beberapa tahapan untuk menangani *missing value, noise, outlier, dan inconsistent data*. Tabel I menunjukkan contoh data transaksi perusahaan sebelum dilakukan proses penanganan *missing value*. Tabel II menunjukkan data yang telah di-*update* dengan mengisi bagian *missing value*. Selanjutnya, dilakukan penanganan *noise* pada data transaksi. Setelah dilakukan analisis, ditemukan adanya *noise* berupa data retur yang dicatat dalam data transaksi. Tabel III menunjukkan data transaksi yang mengandung *noise*. *Noise* dihapus pada data transaksi

sehingga data dapat digunakan dengan lebih maksimal. Tabel IV menunjukkan data transaksi yang telah dibersihkan dari *noise*. Proses *data cleaning* selanjutnya adalah pengecekan *outlier*. Pada penelitian ini, pengecekan *outlier* dilakukan dengan alat bantu berupa aplikasi Excel dan WEKA. Hasil dari pengujian dengan menggunakan kedua alat bantu tersebut adalah data transaksi yang digunakan tidak memiliki nilai *outlier*. Terakhir, dilakukan pengujian konsistensi data terhadap data atribut nama produk dengan menggunakan format “kode kategori + kode produk”. Tabel V menunjukkan data transaksi dengan format nama produk yang telah ditentukan.

Setelah *cleaning data*, dilakukan *data reduction* dengan menggunakan *feature selection*. Fungsi yang digunakan ini bersifat *subjektif* sesuai dengan tujuan dan data transaksi perusahaan. Oleh karena itu, data atribut yang digunakan dalam penelitian ini adalah:

1. Tanggal Transaksi

Data atribut ini memberikan informasi mengenai waktu pembelian setiap pelanggan di perusahaan.

2. Nama Pelanggan

Data atribut ini memberikan informasi mengenai identitas pelanggan yang akan digunakan untuk proses segmentasi pelanggan.

3. Jumlah Produk

Data atribut ini memberikan informasi mengenai kebiasaan pembelian pelanggan dalam bentuk jumlah produk.

4. Total Pembayaran

Data atribut ini memberikan informasi mengenai kebiasaan pembelian pelanggan dalam bentuk total pembayaran.

TABEL I

CONTOH *MISSING VALUE*

Tanggal Transaksi	Nama Pelanggan	Nama Produk	Jumlah Produk	Ukuran Produk	Harga (Rp)	Total (Rp)	Kota
8/29/2019	Sri***	Scarf Pink	1	-	135.000	135.000	Kendal
8/9/2019	Sri***	Scarf Merah	1	-	145.000	145.000	Kendal

TABEL II

CONTOH *UPDATE MISSING VALUE*

Tanggal Transaksi	Nama Pelanggan	Nama Produk	Jumlah Produk	Ukuran Produk	Harga (Rp)	Total (Rp)	Kota
8/29/2019	Sri***	Scarf Pink	1	All Size	135.000	135.000	Kendal
8/9/2019	Sri***	Scarf Merah	1	All Size	145.000	145.000	Kendal

TABEL III

CONTOH *NOISE*

Tanggal Transaksi	Nama Pelanggan	Nama Produk	Jumlah Produk	Ukuran Produk	Harga (Rp)	Total (Rp)	Kota
8/5/2019	Mir***	Gamis GA210	1	XL	485.000	485.000	Surabaya
8/5/2019	Mir***	Blouse B100	1	-	0	0	Kendal

TABEL IV

CONTOH PEMBERSIHAN *NOISE*

Tanggal Transaksi	Nama Pelanggan	Nama Produk	Jumlah Produk	Ukuran Produk	Harga (Rp)	Total (Rp)	Kota
8/29/2019	Mii***	Gamis GA210	1	XL	485.000	485.000	Surabaya

TABEL V

FORMAT DATA TRANSAKSI

Tanggal Transaksi	Nama Pelanggan	Nama Produk	Jumlah Produk	Harga (Rp)	Total (Rp)	Kota
9/16/2019	Erd***	S-3	1	145.000	145.000	Sekupang
9/17/2019	Nad***	B-106	1	375.000	375.000	Tambalang
9/17/2019	San***	B-101	1	375.000	375.000	Batu

### B. Transformasi Data

Proses terakhir yang dilakukan pada tahap *pre-processing* data adalah proses transformasi dengan menggunakan metode RFM. Metode RFM akan menghasilkan 3 data atribut yang mempengaruhi kemungkinan pembelian oleh pelanggan di masa depan dan membantu proses segmentasi pelanggan di CV XYZ sehingga dapat diketahui informasi pengelompokan pelanggan yang dibutuhkan oleh perusahaan. Tiga data atribut yang dihasilkan dalam penggunaan metode RFM ini adalah *recency*, *frequency*, dan *monetary*.

Setelah dilakukan proses pengelompokan dengan menggunakan data atribut *frequency*, diketahui bahwa data atribut ini belum dapat memberikan hasil segmentasi pelanggan yang signifikan. Oleh karena itu, dilakukan penawaran alternatif, yaitu rata-rata pembelian item per bulan. Hasil akhir dari transformasi data yang akan digunakan adalah data atribut *recency*, rata-rata pembelian *item* per bulan, dan *monetary*. Tabel VI menunjukkan data atribut hasil transformasi data yang akan digunakan untuk proses *clustering* pelanggan. Metode *clustering* dapat diterapkan untuk mengelompokkan pelanggan sesuai dengan karakteristik atau profil pelanggan, seperti yang dilakukan di [7] dan [8].

### C. Clustering Data

Proses *clustering* data dilakukan dalam penelitian ini karena data atribut yang dihasilkan pada proses transformasi belum dapat digunakan untuk proses segmentasi pelanggan. Oleh karena itu, metode *clustering* diterapkan untuk mencapai tujuan yang diinginkan oleh perusahaan. Metode *clustering* dipilih dikarenakan metode ini dapat memenuhi kebutuhan dalam penelitian ini, yaitu:

1. Dapat menyimpan dan mengolah data transaksi penjualan CV XYZ.
2. Dapat memilih data atribut yang sesuai dengan tujuan yang ingin dicapai.
3. Dapat menghasilkan *cluster* pelanggan berdasarkan data atribut yang dimiliki.
4. Dapat menghasilkan *cluster* pelanggan berdasarkan kemiripan nilai data atribut di CV XYZ.

TABEL VI

HASIL TRANSFORMASI DATA

Nama Pelanggan	<i>Recency</i>	Rata-rata Pembelian Item/Bulan	<i>Monetary</i> (Rp)
Ade***	64 Hari	1 Transaksi	66.875
Adi***	206 Hari	1 Transaksi	37.250
Afa***	33 Hari	1 Transaksi	71.875

Pada proses *clustering*, pelanggan akan melakukan analisis dan perbandingan untuk mendapatkan algoritme dan jumlah *cluster* yang paling baik dan tepat. Algoritme yang dibandingkan dalam penelitian ini *K-means* dan *hierarchical*. Setelah dilakukan *clustering* dengan menggunakan kedua algoritme tersebut, dilakukan analisis menggunakan metode evaluasi untuk menentukan algoritme mana yang paling baik dan tepat dalam proses *clustering* pelanggan.

Metode evaluasi yang digunakan adalah perhitungan dengan menggunakan simpangan baku yang akan menghasilkan sebuah rasio. Rasio ini yang akan digunakan untuk menentukan algoritme yang paling tepat untuk digunakan. Algoritme yang memiliki nilai rasio terkecil merupakan algoritme yang paling tepat untuk digunakan pada proses *clustering*. Rumus untuk menghitung simpangan baku adalah sebagai berikut [9]:

- a. Rata-rata simpangan baku dalam *cluster* ( $S_w$ )

$$S_w = K^{-1} \sum_{k=1}^k S_k \quad (2)$$

$K$  adalah banyaknya *cluster* yang terbentuk dan  $S_k$  adalah simpangan baku *cluster* ke- $k$ .

- b. Simpangan baku antar *cluster* ( $S_b$ )

$$S_b = \left[ (K-1)^{-1} \sum_{k=1}^k (\bar{X}_k - \bar{X})^2 \right]^{\frac{1}{2}} \quad (3)$$

$\bar{X}_k$  adalah rata-rata *cluster* ke- $k$  dan  $\bar{X}$  adalah rata-rata keseluruhan *cluster*.

Jumlah *cluster* yang dibandingkan dalam penelitian ini adalah:

1. Tiga *cluster*
  - a. *Cluster* 1: Pelanggan dengan tingkat pembelian sangat rendah dan sudah lama tidak melakukan transaksi.
  - b. *Cluster* 2: Pelanggan yang jarang berkunjung, tetapi nilai transaksi rata-rata tinggi.
  - c. *Cluster* 3: Pelanggan yang sering melakukan transaksi dan memiliki nilai transaksi tertinggi.
2. Empat *cluster*
  - a. *Cluster* 1: Pelanggan dengan tingkat pembelian sangat rendah dan sudah lama tidak melakukan transaksi.
  - b. *Cluster* 2: Pelanggan yang jarang berkunjung, tetapi nilai transaksi rata-rata tinggi.
  - c. *Cluster* 3: Pelanggan yang sering melakukan transaksi dan memiliki nilai transaksi tertinggi kedua.
  - d. *Cluster* 4: Pelanggan yang sering melakukan transaksi dan memiliki nilai transaksi tertinggi.
3. Lima *cluster*
  - a. *Cluster* 1: Pelanggan dengan tingkat pembelian sangat rendah dan sudah lama tidak melakukan transaksi.
  - b. *Cluster* 2: Pelanggan yang jarang berkunjung dan nilai transaksi rendah.
  - c. *Cluster* 3: Pelanggan yang jarang berkunjung, tetapi nilai transaksi rata-rata tinggi.
  - d. *Cluster* 4: Pelanggan yang sering melakukan transaksi dan memiliki nilai transaksi tertinggi kedua.
  - e. *Cluster* 5: Pelanggan yang sering melakukan transaksi dan memiliki nilai transaksi tertinggi

Untuk menentukan jumlah *cluster* yang paling tepat dalam penelitian ini, digunakan sebuah metode evaluasi yaitu metode *Davies Bouldin Index (DB Index)*. Metode ini merupakan sebuah metode evaluasi untuk mengukur jumlah kelompok yang terbaik pada suatu metode *clustering*. Pendekatan pengukuran *DB Index* yaitu memaksimalkan jarak antar *cluster* serta meminimalkan jarak intra *cluster*. Selain itu, *DB Index* mengukur rata-rata kemiripan antara masing-masing *cluster* dan salah satu yang paling mirip. Nilai *DB Index* yang minimum adalah skema *clustering* yang paling optimum. Semakin kecil nilai yang didapatkan, semakin baik *clustering* yang dilakukan, begitu juga sebaliknya. Rumus untuk menghitung *DB Index* adalah sebagai berikut [10]:

$$\text{var}(x) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (4)$$

$$R_{ij} = \frac{\text{var}(C_i) + \text{var}(C_j)}{\|C_i - C_j\|}, \quad i \neq j \quad (5)$$

$$R_i = \max R_{ij}, \quad j = 1, \dots, k; \quad i \neq j \quad (6)$$

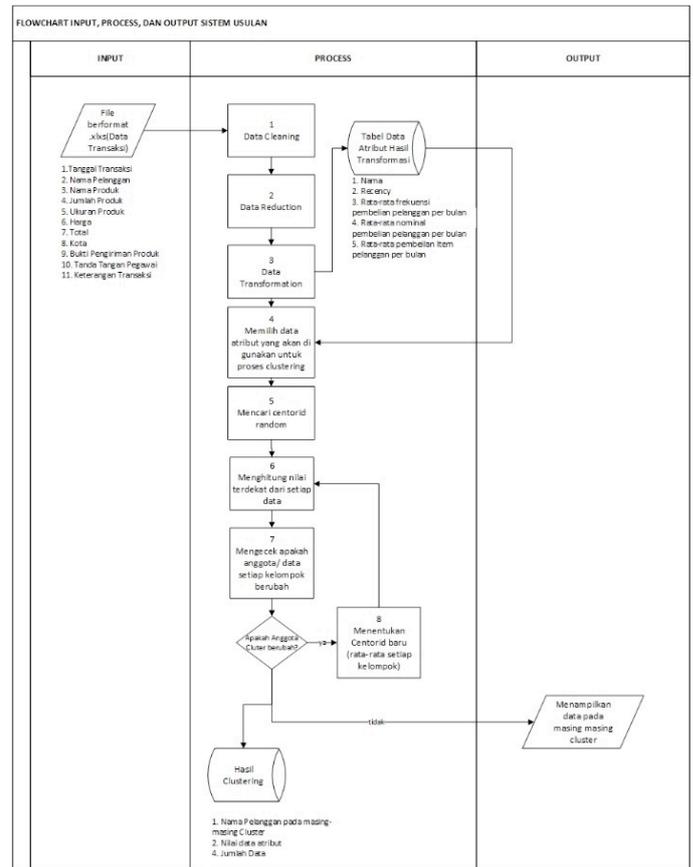
$$\text{DB Index} = \frac{1}{k} \sum_{i=1}^k R_i \quad (7)$$

Empat rumus di atas akan digunakan secara berurutan sehingga didapatkan nilai *DB Index* untuk setiap pengelompokan yang dilakukan. Setiap nilai *DB Index* yang didapatkan akan dibandingkan sehingga diketahui jumlah *clustering* pelanggan yang paling tepat.

Setelah dilakukan analisis dan evaluasi dengan menggunakan metode pengukuran, diperoleh hasil bahwa algoritme *K-Means* dengan *clustering* menggunakan 3 *cluster* merupakan algoritme dan jumlah *cluster* yang paling tepat dalam proses *clustering* pada penelitian ini.

#### D. Implementasi Sistem

Sistem yang dirancang dalam penelitian ini berdasarkan model *clustering* yang dibuat sesuai dengan diagram *input-process-output* dalam Gambar 2. Sistem *clustering* yang diimplementasikan memiliki tampilan awal berupa sebuah tabel dan beberapa data atribut yang dapat digunakan dalam proses *clustering*, seperti tampak pada Gambar 3.



Gambar 2 Skema IPO sistem *clustering*

User pertama-tama harus melakukan *import data* dengan menekan tombol "import data". Selanjutnya, user akan masuk ke halaman berikutnya (Gambar 4).

Setelah masuk ke halaman tersebut, user diberikan beberapa fungsi seperti *download format*, *choose file*, *preview*, dan *cancel*. Berikut merupakan penjelasan dari setiap fungsi tersebut:

1. *Download Format*

Fungsi ini dapat digunakan jika user membutuhkan atau tidak memiliki format data untuk menyimpan data transaksi yang terjadi di CV XYZ. Format tersebut berupa sebuah file Excel dengan format .xlsx dan berisi atribut-atribut sesuai dengan format data transaksi CV XYZ.

2. *Choose File*

Fungsi ini digunakan oleh user untuk melakukan impor data ke dalam sistem. User dapat memilih file yang berisikan data transaksi yang ingin diolah (Gambar 5).

3. *Preview*

Fungsi ini digunakan ketika user sudah berhasil melakukan impor data ke dalam sistem. Sistem akan menampilkan data sesuai dengan isi dari file yang diimpor, seperti tampak pada Gambar 6.

4. *Import*

Fungsi ini digunakan oleh user ketika ingin melakukan impor data untuk diolah di dalam sistem. Ketika file sudah berhasil diimpor, maka sistem akan menyimpan data tersebut pada database. Data tersebut akan otomatis diolah pada tahap *pre-processing* dan ditampilkan pada halaman awal, seperti pada Gambar 7.

Sistem memiliki fungsi untuk melakukan proses transformasi data. Fungsi tersebut digunakan untuk mengolah data yang telah diimpor dan telah melalui proses *pre-process-*

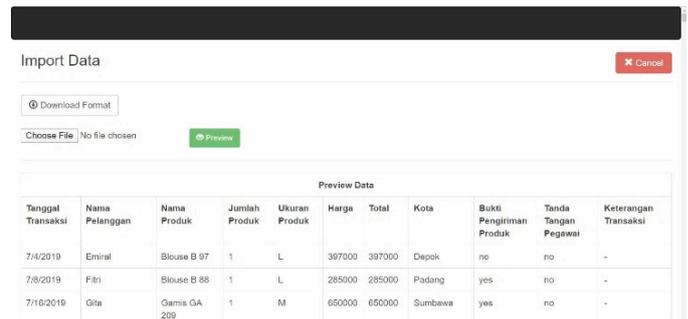
ing menjadi atribut-atribut data yang akan digunakan untuk proses *clustering*. Proses transformasi data dilakukan berdasarkan metode yang telah ditentukan pada sebelumnya, yaitu menggunakan metode RFM. Ketika tombol "Transformasi data" diklik, user akan masuk ke halaman berikutnya yang menampilkan hasil transformasi data (Gambar 8).

Proses transformasi data akan menghasilkan tiga buah atribut, yaitu *recency*, rata-rata pembelian *item*, dan nominal per bulan. Setelah proses transformasi dilakukan, data akan tersimpan dalam database sistem. Dalam halaman transformasi ini terdapat sebuah tombol untuk memilih atribut segmentasi. Tombol tersebut dapat digunakan oleh user ketika akan melakukan proses segmentasi data menggunakan atribut data yang telah terbentuk. Fungsi utama dalam sistem ini adalah untuk melakukan *clustering* pelanggan. Fungsi ini digunakan oleh user ketika ingin melakukan proses *clustering* menggunakan atribut-atribut yang sudah dilakukan proses transformasi. Pertama-tama user akan menekan tombol "Pilih data atribut clustering". Ketika user menekan tombol tersebut, maka sistem akan menampilkan halaman pilih atribut (Gambar 9). Pada halaman tersebut, user akan memilih atribut yang akan digunakan untuk melakukan *clustering*.

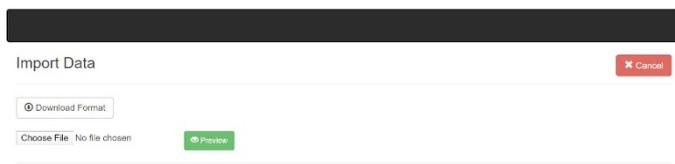
Terdapat tiga pilihan atribut yang dapat digunakan untuk proses segmentasi, yaitu *recency*, rata-rata pembelian *item*, dan nominal pelanggan untuk setiap bulannya. Pada tahap ini user dapat melakukan kombinasi atribut untuk mendapatkan hasil segmentasi atau segmentasi pelanggan yang diinginkan.



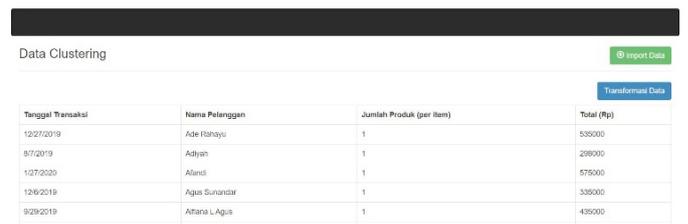
Gambar 3 Halaman awal sistem clustering



Gambar 6 Tampilan preview data



Gambar 4 Halaman import data



Gambar 7 Data import



Gambar 5 Fungsi choose file



Gambar 8 Hasil transformasi data

Ketika *user* telah memilih atribut data yang diinginkan, sistem akan menampilkan halaman berikutnya (Gambar 10).

Data ditampilkan sesuai dengan atribut yang dipilih oleh *user*. Selanjutnya, *user* dapat kembali memilih atau mengganti atribut untuk dilakukan *clustering* pelanggan. Ketika atribut yang dipilih sudah sesuai, *user* dapat memasukan jumlah data kelompok yang yang diinginkan. Jika sudah sesuai, maka *user* dapat menekan tombol “*Clustering data*” lalu sistem akan mengolah data tersebut menggunakan algoritme *K-means*. Setelah proses pengolahan data berhasil dilakukan, sistem akan menampilkan halaman berikutnya (Gambar 11).

Halaman hasil segmentasi pengelompokan pelanggan menampilkan *scatter plot* yang berisi informasi bagaimana penyebaran hasil segmentasi pelanggan berdasarkan atribut yang telah dipilih oleh *user*. Selanjutnya, ditampilkan *pie diagram* yang berisi presentase masing-masing *cluster* pelanggan. Terakhir, terdapat tombol “Detail segmentasi data” yang dapat digunakan untuk menampilkan daftar setiap pelanggan berdasarkan pembagian masing-masing *cluster* dan total anggota setiap *cluster*, seperti pada Gambar 12.

#### IV. SIMPULAN

Pemanfaatan data transaksi dapat dilakukan dengan melakukan segmentasi menggunakan teknik *clustering* untuk mendapatkan pengetahuan mengenai segmentasi pelanggan di CV XYZ.

Penentuan algoritme dan jumlah *cluster* diperoleh dari proses evaluasi menggunakan teknik simpangan baku dan metode *Davies Bouldin Index*. Hasilnya, *K-means* paling tepat untuk diterapkan pada data transaksi CV XYZ dengan membagi pelanggan menjadi 3 *cluster*.

Hasil pengolahan yang dilakukan pada data transaksi milik CV XYZ menghasilkan data atribut *recency*, *monetary*, dan rata-rata pembelian *item* yang dapat digunakan untuk melakukan *clustering* pelanggan dengan menggunakan algoritme *K-means*.

Berdasarkan *clustering* pelanggan dengan menggunakan algoritme *K-means* dan evaluasi pengukuran jumlah kelompok dengan metode DB Index, diperoleh hasil bahwa *clustering* pelanggan dengan menggunakan 3 *cluster* pelanggan merupakan jumlah kelompok yang paling tepat untuk data atribut:

- Rata-rata pembelian *item*/bulan;
- Monetary*;
- Recency* dan rata-rata pembelian *item*/bulan;
- Recency* dan *monetary*;
- Rata-rata pembelian *item*/bulan dan *monetary*;
- Recency*, rata-rata pembelian *item*/bulan, dan *monetary*.

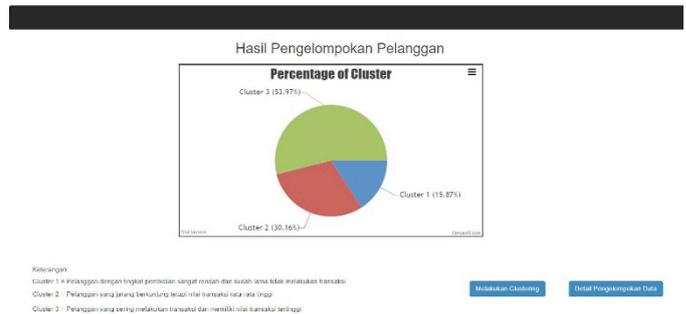
*Clustering* pelanggan dengan menggunakan 4 *cluster* pelanggan merupakan jumlah kelompok yang paling baik untuk data atribut *recency*.

Nama Pelanggan	Recency (Hari)	Rate-rata Pembelian Item Pelanggan (Transaksi per Bulan)	Rate-rata Nominal Pelanggan (Rupiah per Bulan)
Ade Rahayu	64	1	66875
Adyah	206	1	37250
Alvin	33	1	71875
Agus Suardar	66	1	41875
Alana L. Agus	163	1	120250

Gambar 9 Halaman pilih atribut

Nama Pelanggan	Recency (Hari)	Rate-rata Nominal Pelanggan (per Bulan)
Ade Rahayu	64	66875
Adyah	206	37250
Alvin	33	71875
Agus Suardar	66	41875
Alana L. Agus	163	120250
Alina	707	66250

Gambar 10 Hasil pemilihan atribut



Gambar 11 Halaman hasil clustering

Nama Pelanggan C1	Recency (Hari)	Rate-rata Nominal Pelanggan (per Bulan)
Adyah	206	37250
Agus Suardar	66	41875
Alvin	33	71875
Alana L. Agus	163	120250
Alina	707	66250

Gambar 12 Halaman detail clustering

#### DAFTAR REFERENSI

- [1] M. Ferguson, *Big Data- Why Transaction Data is Mission Critical To Success*. England: Intelligent Business Strategies, 2014.
- [2] *Strategic Marketing* (20 Januari, 2019). *Five Benefits of a Customer Database* [Daring]. Tersedia: <https://thinkstrategic.com/five-benefits-of-a-customer-database/>.
- [3] Ayuliana, F. Yosieto. “Pemanfaatan data warehouse proses penjualan dan pembelian untuk dukungan pengambilan keputusan”, *Jurnal Imiah FIFO*, volume VII, no. 2, November 2015.
- [4] A. D. Savitri, F. A. Bachtiar, dan N. Y. Setyawan, “Segmentasi pelanggan menggunakan metode k-means clustering berdasarkan model RFM pada klinik kecantikan (studi kasus: Belle Crown Malang)” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 9, hlm. 2957-2966, September 2018.

- [5] C. C. Aggarwal, *Data Mining: The Textbook*. United States of America: Springer International Publishing, 2015.
- [6] C. C. Aggarwal, *Data Clustering: Algorithms and Application*. United Kingdom: King Taylor & Francis Group, 2014.
- [7] E. M. Sipayung, H. Maharani, dan B. A. Paskhadira, "Designing customer target recommendation system using K-means clustering method", *IJITEE*, vol 1, no. 1, hlm. 1-7, Maret 2017.
- [8] C. Fiarni, H. Maharani, dan N. Calista, "Product Recommendation System Design Using Cosine Similarity and Content-based Filtering Methods", *IJITEE*, vol. 3, no. 2, hlm. 42-48, Juni 2019.
- [9] F. Satria dan R. Z. A. Aziz. "Perbandingan kinerja metode *Ward* dan *K-means* dalam menentukan *cluster* data mahasiswa pemohon beasiswa (studi kasus: STMIK Pringsewu)" *Jurnal TIM Darmajaya* vol. 2, no. 01, hlm. 2442-5567, Mei 2016.
- [10] F. Irhamni, F. Damayanti, B. Khusnul, dan Miffatchul. "Optimalisasi pengelompokan kecamatan berdasarkan indikator pendidikan menggunakan metode *clustering* dan *Davies Bouldin Index*", dalam *Prosiding Seminar Nasional Sains dan Teknologi (Semnastek)* 2014, Jakarta, 12 November 2014.
- Yosi Yonata**, menerima gelar Sarjana Teknik dari ITB Jurusan Teknik Elektro bidang Teknik Komputer pada tahun 2000 dan gelar Magister Teknik dari ITB Jurusan Teknik Elektro bidang Teknologi Informasi pada tahun 2002. Saat ini aktif sebagai dosen tetap di Departemen Sistem Informasi ITHB Bandung.
- Herastia Maharani**, menerima gelar Sarjana Teknik dari Departemen Teknik Informatika Institut Teknologi Bandung (ITB) pada tahun 2005 dan gelar Magister Teknik dari Sekolah Teknik Informatika (STEI) ITB dengan konsentrasi Informatika pada tahun 2010. Saat ini menjabat sebagai dosen tetap di Departemen Sistem Informasi ITHB.
- Chen Viona**, lahir di Pematang Siantar, tahun 1998. Menyelesaikan studi S1 di Program Studi Sistem Informasi Institut Teknologi Harapan Bangsa pada tahun 2020. Minat penelitian pada bidang *data analysis*.