BAB III
ANALISIS DAN PERANCANGAN SISTEM

III.1. Analisis Sistem

Analisis sistem merupakan tahapan untuk mengidentifikasi masalah dan menjalankan serangkaian proses untuk mengatasi masalah pada suatu sistem. Kegiatan pada proses analisis sistem meliputi analisis masalah, penerapan model proses CRISP-DM untuk menyelesaikan masalah *data mining*, dan melakukan analisis kebutuhan sistem yang akan dibangun.

III.1.1. Analisis Masalah

Setiap menghadapi musim kompetisi baru dan *event* tertentu, Jeger Jersey Indonesia selalu melakukan produksi *merchandise* secara besar. Pemasaran produk saat ini dilakukan kepada pelanggan baru dan pelanggan yang sudah ada (*existing customers*). Hal tersebut diharapkan dapat meningkatkan penjualan produk, terutama jenis produk *jersey*. Selain itu, menjaga hubungan baik dengan pelanggan yang sudah ada dapat menjaga loyalitas dan kepuasan pelanggan. Oleh karena itu, manajer *marketing* perlu menerapkan strategi promosi yang tepat terhadap karakteristik pelanggan yang berbeda. Namun semua pelanggan yang sudah ada saat ini masih mendapat perlakuan promosi yang sama, sehingga kurang efektif untuk meningkatkan penjualan produk.

III.1.2. Pemahaman Bisnis

Tahap pertama pada model proses CRISP-DM adalah tahap pemahaman bisnis (*business understanding*). Pada tahap ini, dilakukan pengumpulan informasi mengenai tujuan bisnis dan hasil yang diharapkan oleh manajer *marketing* Jeger Jersey Indonesia. Kemudian, dilakukan penyusunan rencana kegiatan yang akan dilakukan untuk mencapai tujuan yang telah ditentukan. Kegiatan pada tahap ini meliputi identifikasi tujuan bisnis, penilaian situasi, penentuan tujuan *data mining*, dan penyusunan rencana proyek.

a. Identifikasi Tujuan Bisnis

Kegiatan identifikasi tujuan bisnis dilakukan untuk memahami tujuan bisnis dari Jeger Jersey Indonesia secara menyeluruh. Identifikasi juga dilakukan terhadap faktor-faktor penting yang akan memengaruhi hasil akhir penelitian. Berikut merupakan tiga hal penting pada kegiatan identifikasi tujuan bisnis.

1. Latar Belakang

Penjualan *merchandise* tim sepakbola Indonesia selalu meningkat pada momen tertentu, terutama saat Liga Indonesia akan segera dimulai. Jeger Jersey Indonesia sebagai vendor besar selalu melakukan berbagai persiapan menghadapi kesempatan tersebut. Persiapan juga dilakukan oleh manajer *marketing* yang bertanggung jawab untuk memasarkan produk kepada pelanggan. Salah satu kegiatan yang dilakukan manajer *marketing* adalah melakukan analisis pasar. Hal tersebut dilakukan untuk mengetahui karakter pelanggan yang akan menjadi target pemasaran. Target pasar tersebut meliputi pelanggan baru dan pelanggan yang sudah ada (*existing customers*). Pelanggan yang sudah ada perlu diberikan pelayanan khusus untuk menjaga kepuasan dan loyalitas mereka.

2. Tujuan Bisnis

Fokus dari Jeger Jersey Indonesia saat ini adalah melakukan promosi penjualan kepada pelanggan yang sudah ada. Hal tersebut dilakukan untuk meningkatkan penjualan produk, terutama jenis produk *jersey*. Pelanggan yang pernah melakukan transaksi pembelian sebelumnya dapat menjadi pasar yang potensial. Para pelanggan tersebut diharapkan akan melakukan pembelian kembali dan membeli produk *jersey* lebih banyak.

3. Kriteria Sukses Bisnis

Menurut manajer *marketing,* aspek penting yang menjadi kriteria tercapainya tujuan bisnis di atas adalah meningkatnya keuntungan dari penjualan produk *jersey*. Salah satu faktor yang dapat memengaruhinya adalah penerapan teknik promosi yang sesuai dengan karakteristik pelanggan. Implementasi dari hal tersebut adalah dengan memasarkan produk yang tepat kepada pelanggan yang tepat berdasarkan perilaku pembelian mereka.

b. Penilaian Situasi

Pada kegiatan penilaian situasi, dilakukan pencatatan terhadap semua fakta-fakta yang ditemukan di tempat penelitian secara rinci. Fakta tersebut meliputi ketersediaan sumber daya, kebutuhan yang diperlukan, kemungkinan yang akan terjadi, dan keuntungan yang akan didapat dari penelitian *data mining*. Berikut merupakan hasil dari kegiatan penilaian situasi di Jeger Jersey Indonesia.

1. Ketersedian sumber daya

Pada kegiatan ini dicatat berbagai sumber daya yang tersedia untuk kebutuhan penelitian. Sumber daya tersebut meliputi perangkat keras, perangkat lunak, data, dan personel. Sumber daya yang tersedia di Jeger Jersey Indonesia untuk penelitian ini adalah sebagai berikut.

a. Ketersediaan perangkat keras dan perangkat lunak

Tabel III.1 merupakan perangkat keras yang tersedia. Perangkat keras ini akan digunakan untuk menjalankan perangkat lunak yang relevan dengan kebutuhan penelitian.

Tabel III.1 Perangkat keras yang tersedia

Jenis perangkat keras	Spesifikasi
Prosesor	4 *core* dengan kecepatan 1800 MHz
RAM	4 GB
Harddisk	500 GB
Monitor	14 inci dengan resolusi 1366 x 768 piksel
Printer	Fitur tambahan *scan* dan *copy*
Modem	Modem nirkabel

Sedangkan Tabel III.2 merupakan perangkat lunak yang tersedia. Perangkat lunak tersebut akan digunakan untuk menjalankan aplikasi pengolahan *data mining,* membangun sistem hasil analisis, dan membuat laporan penelitian.

Tabel III.2 Perangkat lunak yang tersedia

Jenis perangkat lunak	Spesifikasi
Sistem operasi	Fedora Workstation 21 64 bit
Lingkungan desktop	GNOME 3.14
Office suite	LibreOffice 4.4

b. Ketersediaan Data

Data yang tersedia untuk penelitian adalah data produk dan data pelanggan Jeger Jersey Indonesia. Data tersebut dikumpulkan dan diolah oleh manajer *marketing* sebagai salah satu usaha dalam melakukan analisis pasar. Untuk keperluan penelitian, peneliti telah mendapat izin dari *owner* dan manajer *marketing* untuk mengakses dan mengolah data tersebut.

c. Ketersediaan Personel

Personel yang tersedia berasal dari internal dan eksternal Jeger Jersey Indonesia. Beberapa personel yang menjadi bagian dari penelitian ini adalah sebagai berikut.

1. *Owner* Jeger Jersey Indonesia sebagai sponsor penelitian.
2. Manajer *marketing* Jeger Jersey Indonesia sebagai analis pasar dan pengguna hasil penelitian *data mining*.
3. Peneliti bertugas untuk melakukan penelitian *data mining.*
4. Pembimbing penelitian sebagai *data mining expert* yang mengarahkan peneliti dalam melakukan penelitian

2. Kebutuhan dan asumsi

Hasil penelitian ini dibutuhkan oleh manajer *marketing* untuk mendukung pencapaian tujuan bisnis yang telah ditetapkan. Hasil penelitian diharapkan dapat memenuhi kriteria sukses bisnis. Untuk mencapai hal tersebut, diperlukan data pelanggan yang memiliki atribut-atribut yang mencerminkan perilaku pembelian pelanggan. *Data preprocessing* akan dilakukan terlebih dahulu jika data yang digunakan belum memiliki kualitas yang baik. Peneliti berasumsi bahwa hasil penelitian akan didapat dalam jangka satu bulan setelah penelitian dimulai.

3. Resiko dan kemungkinan

Resiko pada penelitian ini meliputi kemungkinan kejadian yang akan berdampak pada waktu, usaha, dan hasil penelitian. Beberapa resiko dan kemungkinan yang dapat terjadi dalam penelitian ini adalah sebagai berikut.

a. Penggunaaan alat dan metode *data mining* yang berbeda untuk menyesuaikan dengan data yang diperoleh dan hasil yang diharapkan.

b. Waktu yang dibutuhkan untuk *data preprocessing* cukup lama.

c. Hasil penelitian *data mining* tidak sesuai dengan kriteria karena data yang dibutuhkan tidak terpenuhi.

4. Terminologi

Daftar istilah yang digunakan pada tahap pemahaman bisnis dijelaskan pada Tabel III.3. Istilah-istilah berikut terdiri dari istilah pada bisnis dan istilah pada *data mining*.

Tabel III.3 Terminologi pada tahap pemahaman bisnis

No.	Terminologi	Penjelasan
1	*Existing customers*	Pelanggan yang pernah melakukan transaksi pembelian sebelumnya.
2	*Data preprocessing*	Sejumlah proses pengolahan data untuk menghasilkan data yang berkualitas.
3	*Data mining expert*	Seseorang yang ahli dalam disiplin *data mining*.
4	*Office suite*	Kumpulan perangkat lunak untuk kebutuhan perkantoran seperti mengolah angka, *database*, presentasi, dan lain-lain.
5	*Clustering / cluster analysis*	Teknik *data mining* untuk melakukan pengelompokan data.
6	*Open source software*	Perangkat lunak yang secara bebas dapat digunakan, diubah, dan dibagikan (dengan atau tanpa modifikasi) oleh siapa pun [12].

5. Keuntungan

Hasil penelitian *data mining* ini dapat digunakan oleh manajer *marketing* dalam usaha promosi penjualan kepada pelanggan yang sudah ada. Manajer *marketing* dapat mencoba strategi promosi yang berbeda pada berbagai target pelanggan. Hasil dari usaha tersebut dapat diukur dan dibandingkan untuk menemukan promosi paling efektif untuk setiap target pelanggan.

c. Penentuan Tujuan Data Mining

Kegiatan ini bertujuan untuk mendeskripsikan hasil yang diharapkan dari penelitian, sehingga tujuan bisnis dapat tercapai. Pada tahap ini dijelaskan tujuan proyek dari sudut pandang teknik. Berikut merupakan hasil dari kegiatan penentuan tujuan *data mining*.

1. Tujuan Data Mining

Dalam memasarkan produk kepada pelanggan yang sudah ada, perlu diterapkan promosi yang sesuai dengan karakter pelanggan. Manajer *marketing* menyebutkan bahwa perlu dilakukan segmentasi terhadap pelanggan yang sudah ada. Manajer *marketing* dapat menentukan pendekatan promosi yang tepat untuk setiap segmen yang terbentuk. Segmentasi pelanggan tersebut dapat dilakukan menggunakan salah satu teknik *data mining* yaitu metode *clustering*.

2. Kriteria Sukses Data Mining

Kriteria sukses pada penelitian ini adalah menghasilkan pengetahuan berupa terbentuknya segmen pelanggan yang sudah ada. Segmentasi pelanggan yang terbentuk tentunya harus memiliki kualitas yang baik. Karakteristik pelanggan pada sebuah segmen harus dapat dibedakan dengan pelanggan pada segmen lain.

d. Penyusunan Rencana Proyek

Kegiatan ini bertujuan untuk menjelaskan rencana pencapaian tujuan *data mining* sehingga dapat mencapai tujuan bisnis. Berikut merupakan hasil dari kegiatan penyusunan rencana proyek.

1. Rencana Proyek

Beberapa tahapan kegiatan yang akan dilakukan disusun sebagai rencana pencapaian tujuan *data mining*. Setiap kegiatan disertai dengan estimasi waktu yang dibutuhkan untuk menyelesaikannya. Tabel III.4 merupakan estimasi waktu untuk setiap kegiatan yang dilakukan pada penelitian.

Tabel III.4 Estimasi waktu penelitian

No.	Tahapan Kegiatan	Estimasi Waktu (hari)
1	Pemahaman Bisnis	1
2	Pemahaman Data	2
3	Persiapan Data	5
4	Pemodelan	2
5	Analisis kebutuhan sistem	5
6	Perancangan sistem	5
7	Implementasi dan pengujian sistem	20
	Total	40

2. Penentuan Alat dan Teknik Data Mining

Terdapat banyak alat (*tools*) yang dapat digunakan untuk membantu menyelesaikan masalah *data mining.* Dalam penelitian ini, peneliti akan menggunakan perangkat lunak kode sumber terbuka (*open source software*). Penggunaan perangkat lunak *open source* sangat disarankan karena memiliki dokumentasi yang lengkap dan terbuka, mudah untuk dikustomisasi, dan kemudahan berkolaborasi dengan komunitas. Perangkat lunak yang akan digunakan untuk menganalisis data yaitu R [13] dan RapidMiner Studio [14]. Sedangkan sistem segmentasi pelanggan akan dibangun menggunakan bahasa pemrograman Python [15]. Kemudian teknik yang dipakai untuk memenuhi kriteria sukses *data mining* yaitu metode *clustering* atau *cluster analysis.*

III.1.3. Pemahaman Data

Tahap kedua pada model proses CRISP-DM adalah tahap pemahaman data (*data understanding*). Pada tahap ini, dilakukan pengumpulan dan analisis terhadap data yang diperoleh dari Jeger Jersey Indonesia. Kegiatan pada tahap ini meliputi pengumpulan data awal, pendeskripsian data, eksplorasi data, dan verifikasi kualitas data.

a. Pengumpulan Data Awal

Data yang diperoleh adalah data pelanggan yang melakukan pembelian pada tahun 2013-2014. Data tersebut merupakan data pelanggan yang membeli berbagai produk seperti *jersey,* tas, jaket, dan lain-lain. Namun data yang akan digunakan hanya data pelanggan yang melakukan pembelian produk *jersey*. Hal tersebut disesuaikan dengan tujuan bisnis yaitu meningkatkan penjualan produk *jersey.* Data tersebut diperoleh dari manajer *marketing* atas izin dari *owner* Jeger Jersey Indonesia.

b. Pendeskripsian Data

Data pelanggan yang diperoleh disimpan pada sebuah berkas berekstensi *.csv (*comma-separated values)*. Data tersebut memiliki 12 atribut yang terdiri dari 6 atribut identitas pelanggan dan 6 atribut mengenai perilaku pembelian pelanggan. Jumlah baris atau *record* pada data tersebut sebanyak 808 baris.

Tabel III.5 dan Tabel III.6 menyajikan data pelanggan Jeger Jersey Indonesia tahun 2013-2014. Pada tabel tersebut disajikan 10 *record* sebagai contoh.

Tabel III.5 Data pelanggan (atribut ke-1 sampai ke-6)

ID	Nama	Alamat	Kota	Provinsi	Telepon
JJ-0001	ALDE AFRIYOS	Toko *xxx* Tanah Abang	JAKARTA	DKI JAKARTA	081221468 *xxx*
JJ-0005	DHANI PRIANTORO	Jl. Wates *xxx*	BANTUL	D.I. YOGYAKARTA	085729313 *xxx*
JJ-0007	JUSTUS B MARA	Asrama Yonif *xxx*	MALANG	JAWA TIMUR	08125275 *xxx*
JJ-0008	UKY FERDIANSYAH	Jln Veteran *xxx*	LEBONG	BENGKULU	-
JJ-0011	DWI WARDANI	Jl. Telesonic Km. *xxx*	TANGERANG	BANTEN	08561362 *xxx*
JJ-0014	RIZAL PASARRIN	Alamat:Jl,Telkomas,Palapa 6 *xxx*	MAKASSAR	SULAWESI SELATAN	081240726 *xxx*
JJ-0026	FITRIA SOLIHAT	Cigadung Raya Timur *xxx*	BANDUNG	JAWA BARAT	083824553 *xxx*
JJ-0024	RIYADIN NAZAR	Jl R Soeprapto *xxx*	SUMBA TIMUR	NTT	085253337 *xxx*
JJ-0119	MUHAMMAD NABIL ALFAROBI	Jalan kolonel makmun rasyid *xxx*	LAMPUNG SELATAN	LAMPUNG	087838810 *xxx*
JJ-0181	DONA NOVIANA	kp. Cikupa *xxx*	CIAMIS	JAWA BARAT	085223613 *xxx*

Keterangan: *xxx* pada atribut *Alamat* dan *Telepon* digunakan untuk melindungi privasi pelanggan Jeger Jersey Indonesia.

Tabel III.6 Data pelanggan (atribut ke-7 sampai ke-12)

Item	Custom name?	Jumlah item	Jumlah transaksi	Total transaksi	Metode pembayaran
Jersey Persib – XL (sergio #10)	1	1	1	165000	MANDIRI
Persisam Home L, Persipura Home L, Persija Home M, Persija home XL, persija home m	0	5	3	575000	BCA
Jersey Persipura Putih – (J.B.MARA #04)	1	1	1	165000	MANDIRI
Sriwijaya FC Home – M (NIZAM #18), Sriwijaya Away hitam #D.Cobar No: 85 #:y.zand No:13, sriwijaya away putih #nizam N0 17 Size M persipura h0me #DECOBAR N0 23 SIZE M.	1	5	4	1081000	BCA
Mitra Kukar Home - M (Akbar #25)	1	1	1	176000	BCA
Jersey Home Timnas Indonesia - L	0	1	1	125000	MANDIRI
Jersey Home Persija L (#14 Coecoe) dan Jersey Away Arema - L	1	2	1	270000	CASH
Persipura Putih - XL dan L, Ukuran nya m persija dan L arema	1	4	2	542000	BCA
timnas home L persipura home ukuran XL, persija home S, persib home bjb ukuran m2, persib home bjb ukuran XL 1 L1, persija home ukuran XL, persija 3rd ukuran L, persipura away ukuran L	0	10	8	1205000	BNI
-	-	1	1	150000	BCA

Sedangkan penjelasan untuk masing-masing atribut data pelanggan disajikan pada Tabel III.7. Penjelasan meliputi tipe data atribut, nilai atribut dan keterangan.

Tabel III.7 Penjelasan atribut data pelanggan

No.	Artibut	Tipe Data	Nilai	Keterangan
1	ID	String	ASCII	Identitas unik setiap pelanggan.
2	Nama	String	ASCII	Nama pelanggan yang melakukan transaksi.
3	Alamat	String	ASCII	Alamat pelanggan yang melakukan transaksi.
4	Kota	String	ASCII	Kota pelanggan yang melakukan transaksi.
5	Provinsi	String	ASCII	Provinsi pelanggan yang melakukan transaksi.
6	Telepon	String	ASCII	Nomor telepon pelanggan yang melakukan transaksi.
7	Item	String	ASCII	Jenis-jenis produk yang dibeli oleh pelanggan.
8	Custom Name?	Boolean	0:Tidak, 1:Ya	Apakah pelanggan membeli *custom name*?
9	Jumlah Item	Numerik	[0-9]	Jumlah produk yang dibeli oleh pelanggan.
10	Jumlah Transaksi	Numerik	[0-9]	Jumlah transaksi yang dilakukan oleh pelanggan
11	Total Transaksi	Numerik	[0-9]	Total transaksi seluruhnya yang dibayar oleh pelanggan.
12	Metode Pembayaran	String	ASCII	Metode pembayaran yang digunakan oleh pelanggan.

c. Eksplorasi Data

Pada tahap ini, dilakukan analisis secara mendalam terhadap data yang telah diperoleh. Kegiatan ini dilakukan untuk mengetahui karakteristik dan kualitas data tersebut. Hasil dari tahap eksplorasi data dapat menjadi acuan untuk proses selanjutnya yaitu proses persiapan data. Berikut merupakan hasil analisis pada data yang akan digunakan untuk proses *data mining.*

1. Atribut *item*

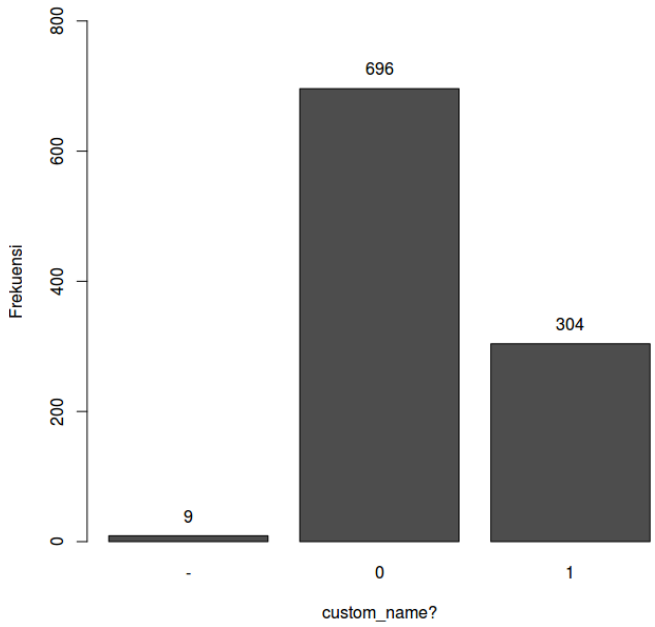
Berdasarkan data pelanggan pada Tabel III.6, atribut *item* mengandung data teks bebas (*free text entries)* yang merepresentasikan jenis *jersey* yang dibeli oleh pelanggan. Selain itu, pada atribut *item* masih terdapat *missing value* yang berpengaruh terhadap atribut lain seperti *custom_name* dan *jumlah_item*. Daftar produk *jersey* yang terdapat pada atribut *item* disajikan pada Tabel III.8.

Tabel III.8 Daftar produk *jersey*

No	Daftar Produk
1	Jersey Persib
2	Jersey PBR
3	Jersey Persija
4	Jersey Semen Padang
5	Jersey Sriwijaya
6	Jersey Arema
7	Jersey PSM
8	Jersey Persipura
9	Jersey Persiba
10	Jersey Persisam
11	Jersey Mitra Kukar
12	Jersey Barito Putera
13	Jersey Timnas

2. Atribut *custom_name*

Atribut *custom_name* merupakan atribut yang memiliki tipe data biner. Nilai 1 merepresentasikan pelanggan yang membeli *custom name,* dan nilai 0 merepresentasikan sebaliknya. Gambar III.1 merupakan *bar chart* yang menggambarkan distribusi nilai dari atribut *custom_name.*



Gambar III.1 Distribusi frekuensi atribut *custom_name*

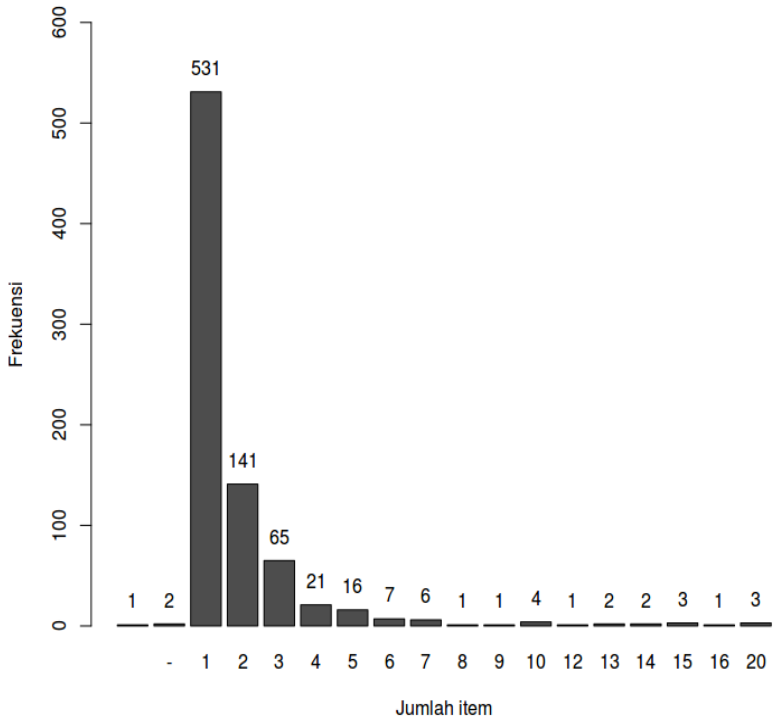
Hasil analisis pada atribut *custom_name* berdasarkan Gambar III.1 adalah sebagai berikut.

a. Atribut ini masih memiliki *missing value* yang ditandai dengan nilai "-".

b. Nilai yang paling banyak muncul pada atribut *custom_name* adalah nilai 0. Hal tersebut menandakan bahwa lebih banyak pelanggan yang hanya memesan *jersey* saja dibandingkan dengan membeli *jersey* dan *custom name.*

3. Atribut *jumlah_item*

Atribut *jumlah_item* berisi jumlah produk yang dibeli oleh pelanggan. Gambar III.2 merupakan *bar chart* yang menggambarkan distribusi nilai dari atribut *jumlah_item.*



Gambar III.2 Distribusi frekuensi atribut *jumlah_item*

Hasil analisis pada atribut *jumlah_item* berdasarkan Gambar III.2 adalah sebagai berikut.

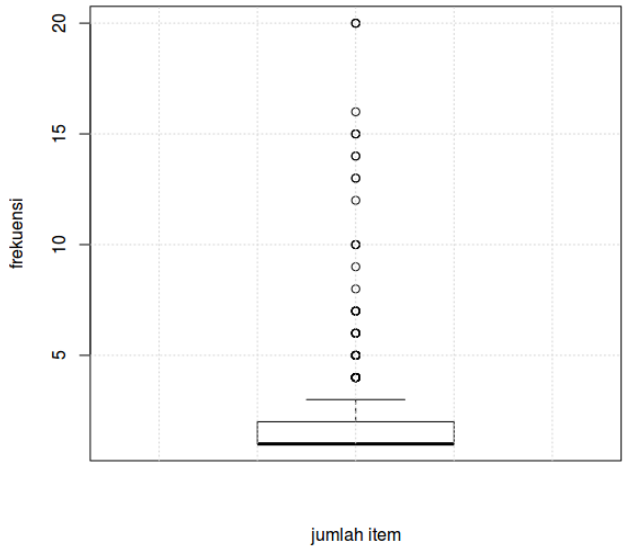
a. Atribut *jumlah_item* masih memiliki *missing value* yang ditandai dengan nilai "-" dan nilai kosong.

b. Nilai yang paling banyak muncul pada atribut ini adalah nilai 1. Hal tersebut

menggambarkan bahwa jumlah pelanggan yang membeli satu buah *jersey* adalah yang paling banyak.

c. Nilai terkecil pada atribut ini adalah 1, sedangkan nilai terbesar adalah 20.

d. Distribusi data pada atribut *jumlah_item* miring ke kanan, sehingga memiliki potensi untuk terdapat *outliers*. Nilai yang bersifat *outliers* perlu diidentifikasi karena akan mempengaruhi model yang dihasilkan. Oleh karena itu, peneliti melakukan perhitungan IQR (*interquartile range*) untuk mengidentifikasi *outliers* dan menyajikan hasilnya dalam *box and whisker plot* pada Gambar III.3.

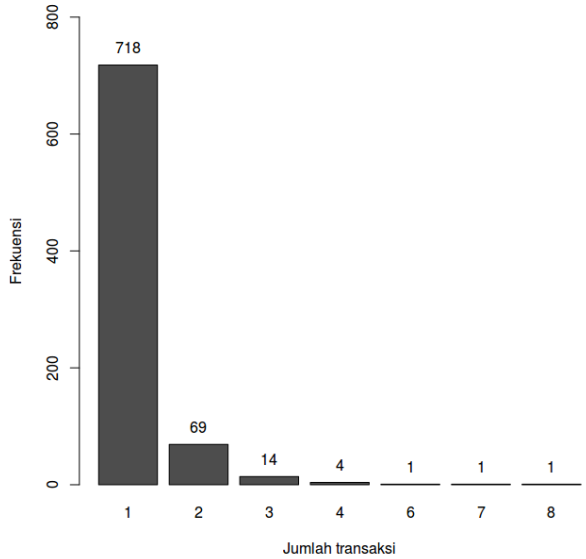


Gambar III.3 *Box and whisker plot* atribut *jumlah_item*

Berdasarkan Gambar III.3, nilai *outliers* merupakan nilai yang berada diatas *whisker*. Pada atribut *jumlah_item* terdapat *outliers* yaitu pelanggan yang membeli lebih dari 3 item. Nilai *outliers* tersebut perlu ditangani dengan menghapusnya atau dilakukan *smoothing* data.

4. Atribut *jumlah_transaksi*

Atribut *jumlah_transaksi* berisi jumlah transaksi pembelian yang dilakukan pelanggan di Jeger Jersey Indonesia. Gambar III.4 merupakan *bar chart* yang menggambarkan distribusi nilai dari atribut *jumlah_transaksi.*

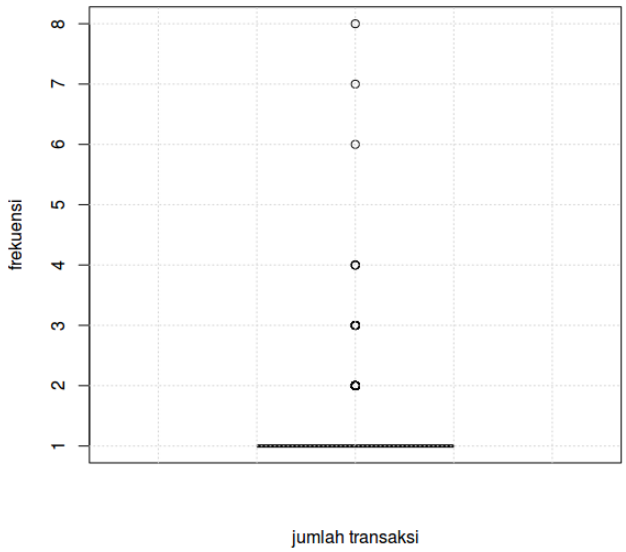


Gambar III.4 Distribusi frekuensi atribut *jumlah_transaksi*

Hasil analisis pada atribut *jumlah_transaksi* berdasarkan Gambar III.4 adalah sebagai berikut.

a. Nilai terkecil pada atribut ini adalah 1, sedangkan nilai terbesar adalah 8.

b. Sama halnya dengan atribut *jumlah_item,* distribusi data pada atribut *jumlah_transaksi* miring ke kanan dan memiliki potensi untuk terdapat *outliers.* Gambar III.5 merupakan *box and whisker plot* yang digunakan untuk mengidentifikasi *outliers* pada atribut *jumlah_transaksi*.



Gambar III.5 *Box and whisker plot* atribut *jumlah_transaksi*

Berdasarkan Gambar III.5, nilai *outliers* merupakan nilai yang berada diatas *whisker*. Pada atribut *jumlah_transaksi* terdapat *outliers* yaitu pelanggan yang melakukan transaksi lebih dari satu kali.

d. Verifikasi Kualitas Data

Pada tahap ini dilakukan pemeriksaan terhadap kualitas data yang diperoleh. Pemeriksaan tersebut meliputi kelengkapan data, validasi data, dan masalah-masalah lain yang terdapat pada data. Berikut hasil dari kegiatan verifikasi kualitas data.

1. Atribut *item* mengandung data teks bebas (*free text entries*). Agar dapat digunakan untuk proses *clustering,* maka atribut ini perlu dikodekan terlebih dahulu ke dalam data numerik.
2. Pada atribut *item*, banyak terdapat nilai yang berbeda tetapi memiliki makna yang sama (*inconsistent*). Sebagai contoh, "jersey persib home" memiliki makna yang sama dengan "jersey persib bandung".
3. Terdapat *missing value* pada atribut *item*, *custom_name,* dan *jumlah_item*.
4. Atribut *jumlah_item* dan *jumlah_transaksi* memiliki nilai yang bersifat *outliers.*

III.1.4. Persiapan Data

Tahap ketiga pada model proses CRISP-DM adalah tahap persiapan data (*data preparation*). Pada tahap ini akan dilakukan pengolahan terhadap data yang telah diperoleh (*data preprocessing*). *Data preprocessing* perlu dilakukan untuk meningkatkan kualitas data agar dapat digunakan pada proses pemodelan. Kegiatan pada tahap ini meliputi pemilihan data, pembersihan data, pembangunan data, dan penyusunan format data.

a. Pemilihan Data

Kegiatan pemilihan data dilakukan untuk menentukan data yang akan digunakan pada tahap pemodelan. Pemilihan data tersebut meliputi pemilihan atribut dan *record* yang akan digunakan. Atribut yang dipilih adalah atribut *ID, item*, *custom name, jumlah_item*, dan *jumlah_transaksi*. Sedangkan *record* yang dipilih adalah semua pelanggan yang membeli produk *jersey.* Atribut dan *record*

tersebut dipilih sesuai dengan tujuan bisnis dari Jeger Jersey Indonesia. Tabel III.9 merupakan contoh data yang dipilih untuk proses pemodelan.

Tabel III.9 Data yang dipilih untuk pemodelan

ID	item	custom_ name	jumlah_ item	jumlah_ transaksi
JJ-0001	Jersey Persib – XL (sergio #10)	1	1	1
JJ-0005	Persisam Home L, Persipura Home L, Persija Home M, persija home XL, persija home m	0	5	3
JJ-0007	Jersey Persipura Putih – (J.B.MARA #04)	1	1	1
JJ-0008	Sriwijaya FC Home – M (NIZAM #18), Sriwijaya Away Hitam #D.Cobar No: 85 #:y.zand No:13, sriwijaya away putih #nizam N0 17 Size M persipura h0me #DECOBAR N0 23 SIZE M.	1	5	4
JJ-0011	Mitra Kukar Home - M (Akbar #25)	1	1	1
JJ-0014	Jersey Home Timnas Indonesia - L	0	1	1
JJ-0026	Jersey Home Persija L (#14 Coecoe) dan Jersey Away Arema - L	1	2	1
JJ-0024	Persipura Putih - XL dan L, Ukuran nya m persija dan L arema	1	4	2
JJ-0119	timnas home L persipura home ukuran XL, persija home S, persib home bjb ukuran m2, persib home bjb ukuran XL 1 L1, persija home ukuran XL, persija 3rd ukuran L, persipura away ukuran L	0	10	8
	-	-	1	1

b. Pembersihan Data

Proses pembersihan data dilakukan untuk meningkatkan kualitas data yang diperoleh. Pembersihan data ini mengacu pada verifikasi kualitas data yang sebelumnya telah dilakukan. Beberapa hal yang dilakukan pada proses pembersihan data adalah sebagai berikut.

1. Menangani data yang tidak konsisten dan *missing value*

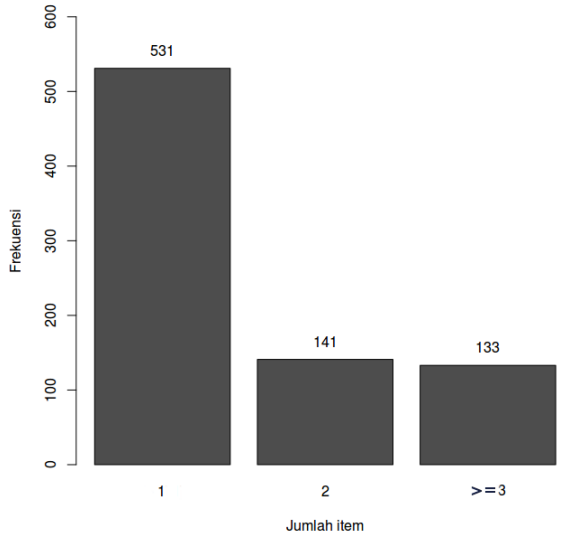
Atribut *item* yang mengandung data teks bebas (*free text entries*) akan dikodekan atau dikonversikan ke dalam data numerik. Setiap jenis produk *jersey* memiliki kode yang berbeda seperti dijelaskan pada Tabel III.8. Hal tersebut dilakukan agar data pada atribut *item* menjadi konsisten. Kemudian *record* yang memiliki *missing value* pada atribut *item* akan dihapus, karena atribut *item* berpengaruh pada atribut *custom_name* dan *jumlah_item*.

2. Menangani *outliers*

Outliers pada atribut *jumlah_item* dan *jumlah_transaksi* akan tetap diproses karena data *outliers* tersebut merupakan data yang valid. Proses *smoothing* akan dilakukan terhadap data pada atribut-atribut tersebut. Salah satu metode yang dapat digunakan untuk *smoothing* data adalah metode *binning*. *Binning* merupakan metode yang digunakan untuk membagi sekumpulan nilai numerik ke dalam beberapa partisi (*bin*). Dengan menggunakan teknik ini, setiap nilai pada atribut akan didistribusikan ke dalam beberapa *bin* yang sudah ditentukan.

a. *Binning* pada atribut *jumlah_item*

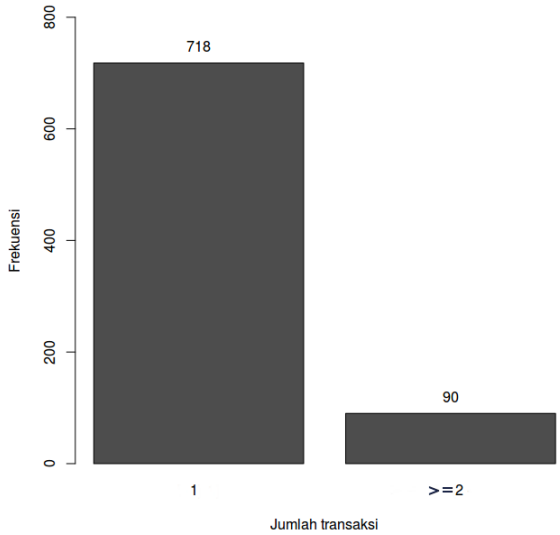
Binning pada atribut *jumlah_item* dimulai dengan mengurutkan setiap nilai terlebih dahulu. Kemudian setiap nilai dipartisi ke dalam *bin* yang kurang lebih memiliki frekuensi yang sama (*equal-frequency partitioning*). Pada atribut ini, jumlah *bin* yang ditentukan adalah 3 *bin*. Gambar III.6 merupakan hasil *binning* pada atribut *jumlah_item.*



Gambar III.6 Hasil *binning* pada atribut *jumlah_item*

b. *Binning* pada atribut *jumlah_transaksi*

Sama seperti pada atribut *jumlah_item,* setiap nilai atribut *jumlah_transaksi* dipartisi ke dalam *bin* yang kurang lebih memiliki frekuensi yang sama (*equal-frequency partitioning*). Pada atribut ini, jumlah *bin* yang ditentukan adalah 2 *bin*. Gambar III.7 merupakan hasil *binning* pada atribut *jumlah_transaksi.*



Gambar III.7 Hasil *binning* pada atribut *jumlah_transaksi*

Tabel III.10 merupakan hasil dari tahap pembersihan data. Data tersebut telah bebas dari nilai yang tidak konsisten, *missing value,* dan *outliers.*

Tabel III.10 Hasil pembersihan data

ID	item	custom_name	jumlah_item	jumlah_transaksi
JJ-0001	1	1	1	1
JJ-0005	10, 8, 3	0	>=3	>=2
JJ-0007	8	1	1	1
JJ-0008	5, 8	1	>=3	>=2
JJ-0011	11	1	1	1
JJ-0014	13	0	1	1
JJ-0026	3, 6	1	2	1
JJ-0024	8, 3, 6	1	>=3	>=2
JJ-0119	13, 8, 3, 1	0	>=3	>=2

c. Pembangunan Data

Tahap pembangunan data digunakan untuk membentuk atribut baru (*attribute construction*). Beberapa atribut seperti *item, jumlah_item,* dan *jumlah_transaksi* perlu diturunkan menjadi beberapa atribut turunan dengan tipe data biner asimetris. Hal tersebut dilakukan agar semua atribut yang digunakan untuk pemodelan memiliki tipe data yang sama, sehingga akan memudahkan perhitungan jarak antar pelanggan.

1. Menurunkan atribut *item*

Setelah dilakukan konversi nilai, atribut *item* menjadi atribut nominal dengan 13 nilai yang merepresentasikan jenis-jenis produk *jersey.* Agar dapat digunakan untuk proses pemodelan, atribut *item* perlu diturunkan menjadi 13 atribut bertipe data biner asimetris. Untuk setiap nilai *v* buat sebuah atribut biner *item_v,* isi dengan nilai 1 jika *item* = *v* dan isi dengan nilai 0 jika sebaliknya. Tabel III.11 merupakan atribut-atribut turunan dari atribut *item.*

Tabel III.11 Hasil penurunan atribut *item*

ID	Item_Persib	Item_PBR	Item_Persija	Item_SPadang	Item_Sriwijaya	Item_Arema	Item_PSM	Item_Persipura	Item_Persiba	Item_Persisam	Item_Kukar	Item_Barito	Item_Timnas
JJ-0001	1	0	0	0	0	0	0	0	0	0	0	0	0
JJ-0005	0	0	1	0	0	0	0	1	0	1	0	0	0
JJ-0007	0	0	0	0	0	0	0	1	0	0	0	0	0
JJ-0008	0	0	0	0	1	0	0	1	0	0	0	0	0
JJ-0011	0	0	0	0	0	0	0	0	0	0	1	0	0
JJ-0014	0	0	0	0	0	0	0	0	0	0	0	0	1
JJ-0026	0	0	1	0	0	1	0	0	0	0	0	0	0
JJ-0024	0	0	1	0	0	1	0	1	0	0	0	0	0
JJ-0119	1	0	1	0	0	0	0	1	0	0	0	0	1

2. Menurunkan atribut *jumlah_item*

Setelah dilakukan *binning,* atribut *jumlah_item* menjadi atribut nominal yang memiliki tiga nilai yaitu 1, 2, dan >=3. Agar dapat digunakan untuk proses pemodelan, atribut ini perlu diturunkan menjadi tiga atribut bertipe data biner asimetris. Ketiga atribut turunan tersebut adalah *jumlah_item_1*, *jumlah_item_2*, dan *jumlah_item_lb3*. Untuk setiap nilai *v*, isi *jumlah_item_v* dengan nilai 1 jika *jumlah_item* = *v* dan isi dengan nilai 0 jika sebaliknya. Tabel III.12 merupakan atribut-atribut turunan dari atribut *jumlah_item.*

Tabel III.12 Hasil penurunan atribut *jumlah_item*

ID	jumlah_item_1	jumlah_item_2	jumlah_item_lb3
JJ-0001	1	0	0
JJ-0005	0	0	1
JJ-0007	1	0	0
JJ-0008	0	0	1
JJ-0011	1	0	0
JJ-0014	1	0	0
JJ-0026	0	1	0
JJ-0024	0	0	1
JJ-0119	0	0	1

3. Menurunkan atribut *jumlah_transaksi*

Setelah dilakukan *binning,* atribut *jumlah_transaksi* menjadi atribut nominal yang memiliki dua nilai yaitu 1 dan >=2. Agar dapat digunakan untuk proses pemodelan, atribut ini perlu diturunkan menjadi dua atribut bertipe data biner asimetris. Kedua atribut turunan tersebut adalah *jumlah_transaksi_1* dan *jumlah_transaksi_lb2.* Untuk setiap nilai *v*, isi *jumlah_transaksi_v* dengan nilai 1 jika *jumlah_transaksi = v* dan isi dengan nilai 0 jika sebaliknya. Tabel III.13 merupakan atribut-atribut turunan dari atribut *jumlah_transaksi.*

Tabel III.13 Hasil penurunan atribut *jumlah_transaksi*

ID	jumlah_transaksi_1	jumlah_transaksi_lb2
JJ-0001	1	0
JJ-0005	0	1
JJ-0007	1	0
JJ-0008	0	1
JJ-0011	1	0
JJ-0014	1	0
JJ-0026	1	0
JJ-0024	0	1
JJ-0119	0	1

d. Penyusunan Data

Proses penyusunan data merupakan langkah terakhir dari tahap persiapan data. Pada tahap ini, dilakukan perubahan sintaksis pada data yang akan digunakan pada saat pemodelan. Selain itu, urutan dari atribut juga harus diperhatikan. Atribut *ID* digunakan sebagai tanda pengenal unik untuk setiap pelanggan, sedangkan 19 atribut selanjutnya merupakan atribut yang akan digunakan untuk pemodelan. Jumlah *record* yang digunakan sebanyak 773 pelanggan. Tabel III.14 menyajikan contoh data yang telah disusun dan siap digunakan untuk pemodelan.

Tabel III.14 Data yang akan digunakan untuk pemodelan

ID	Item_Persib	Item_PBR	Item_Persija	Item_SPadang	Item_Sriwijaya	Item_Arema	Item_PSM	Item_Persipura	Item_Persiba	Item_Persisam	Item_Kukar	Item_Barito	Item_Timnas	custom_name	jumlah_item_1	jumlah_item_2	jumlah_item_lb3	jumlah_transaksi_1	jumlah_transaksi_lb2
JJ-0001	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0
JJ-0005	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	1
JJ-0007	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	1	0
JJ-0008	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	1	0	1
JJ-0011	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	1	0
JJ-0014	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0
JJ-0026	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0	1	0	1	0
JJ-0024	0	0	1	0	0	1	0	1	0	0	0	0	0	1	0	0	1	0	1
JJ-0119	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	1

III.1.5. Pemodelan

Tahap keempat pada model proses CRISP-DM adalah tahap pemodelan (*modeling*). Tahap pemodelan merupakan kegiatan utama dari proses penyelesaian masalah *data mining.* Data yang sudah disusun akan diproses menggunakan suatu metode sehingga dapat menghasilkan model. Kegiatan pada tahap pemodelan meliputi pemilihan teknik pemodelan, penjelasan prosedur pemodelan, penggunaan teknik pemodelan, dan penilaian model.

a. Pemilihan Metode Pemodelan

Langkah pertama pada tahap pemodelan adalah memilih metode yang akan digunakan. Pemilihan metode disesuaikan dengan data telah disiapkan. Metode yang akan digunakan adalah metode *hierarchical agglomerative clustering.* Metode tersebut digunakan karena dapat membentuk segmen-segmen pelanggan secara alami berdasarkan kemiripan nilai atribut-atribut pada data. Selain itu, jumlah segmen pelanggan yang akan dibentuk belum diketahui. Dengan menggunakan metode tersebut, jumlah segmen yang diinginkan dapat ditentukan setelah model terbentuk. Hasil segmentasi pelanggan menggunakan metode tersebut akan disajikan pada sebuah *dendrogram.*

b. Penjelasan Prosedur Pemodelan

Setelah metode pemodelan ditentukan, kemudian akan dijelaskan prosedur dari metode tersebut. Langkah-langkah dari metode *hierarchical agglomerative clustering* adalah sebagai berikut.

1. Merepresentasikan setiap pelanggan sebagai satu kelompok, kemudian menghitung jarak (*distance measure*) antar kelompok-kelompok tersebut.
2. Menggabungkan dua kelompok yang memiliki jarak terdekat.
3. Menghitung jarak antara kelompok yang baru dengan kelompok-kelompok lainnya. Metode yang digunakan adalah metode *complete linkage.*
4. Ulangi langkah kedua dan ketiga sampai tersisa satu kelompok (kumpulan semua pelanggan) dan menyajikan hasil pengelompokkan pada sebuah *dendrogram.*
5. Menentukan jumlah segmen yang diinginkan dengan cara memotong *dendrogram* pada jarak tertentu.

c. Penggunaan Teknik Pemodelan

Pada tahap ini, peneliti akan menerapkan metode pemodelan yang telah dijelaskan sebelumnya. Berikut merupakan proses penerapan metode *hierarchical agglomerative clustering* pada data yang telah disiapkan.

1. Langkah ke-1

Setiap pelanggan direpresentasikan sebagai satu kelompok. Pada data yang telah disiapkan terdapat sebanyak 773 pelanggan, sehingga akan terbentuk 773 kelompok. Setelah itu, dilakukan perhitungan jarak antara setiap kelompok. Metode yang digunakan untuk menghitungnya adalah metode *jaccard distance.* Penggunaan metode tersebut didasarkan pada data yang memiliki tipe data biner asimetris. Sebelum melakukan perhitungan jarak menggunakan metode *jaccard distance*, perlu dibentuk tabel kemungkinan nilai antara dua kelompok yang dibandingkan.

Berikut merupakan contoh perhitungan jarak antara dua kelompok **JJ-0001** dan **JJ-0005**. Berdasarkan data pada Tabel III.14, diketahui nilai-nilai atribut **JJ0001**={1,0,0,0,0,0,0,0,0,0,0,0,0,1,1,0,0,1,0} dan nilai-nilai atribut **JJ0005**={0,0,1,0,0,0,0,1,0,1,0,0,0,0,0,0,1,0,1}. Tabel III.15 merupakan tabel kemungkinan nilai antara kelompok **JJ-0001** dan **JJ-0005**.

Tabel III.15 Tabel kemungkinan nilai antara JJ-0001 dan JJ-0005

JJ-0001 \ JJ-0005	1	0
1	0	4
0	4	10

Langkah selanjutnya adalah melakukan substitusi nilai pada Tabel III.15 ke dalam persamaan *jaccard distance* seperti pada Persamaan II.1. Berikut merupakan perhitungan jarak antara kelompok **JJ-0001** dan **JJ-0005**.

$$dist(JJ\,0001, JJ\,0005) = \frac{n_{10} + n_{01}}{n_{11} + n_{10} + n_{01}}$$

$$dist(JJ\,0001, JJ\,0005) = \frac{4+4}{0+4+4}$$

$$dist(JJ\,0001, JJ\,0005) = \frac{8}{8}$$

$$dist(JJ\,0001, JJ\,0005) = 1$$

Hasil perhitungan jarak antara **JJ-0001** dan **JJ-0005** adalah 1. Nilai 1 menggambarkan bahwa kedua pelanggan tersebut tidak memiliki kemiripan sama sekali. Tabel III.16 menyajikan matriks hasil perhitungan jarak antar setiap kelompok menggunakan metode *jaccard distance.*

Tabel III.16 Matrik jarak antar kelompok pelanggan

	JJ-0001	**JJ-0005**	**JJ-0006**	**JJ-0007**	**JJ-0008**	**JJ-0009**	**JJ-0010**	**JJ-0011**	**JJ-0012**	**JJ-0014**	**. . .**	**JJ-0974**
JJ-0001	0											
JJ-0005	1	0										
JJ-0006	0.25	1	0									
JJ-0007	0.6	0.875	0.4	0								
JJ-0008	1	0.714	0.857	0.857	0							
JJ-0009	0.5	1	0.6	0.6	1	0						
JJ-0010	0.857	0.75	0.714	0.875	0.2	1	0					
JJ-0011	0.6	1	0.4	0.4	0.857	0.6	0.875	0				
JJ-0012	0.5	0.875	0.6	0.6	1	0.5	1	0.6	0			
JJ-0015	0.5	0.857	0.6	0.25	1	0.5	1	0.6	0.5	0		
. **.** **.**	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	
JJ-0974	0.8	0.666	0.833	0.833	0.833	0.8	0.857	0.833	0.5	0.8	. . .	0

2. Langkah ke-2

Setelah jarak antar kelompok pelanggan dihitung, langkah berikutnya adalah menggabungkan dua kelompok yang memiliki jarak terkecil (paling mirip). Berdasarkan matrik jarak antar kelompok pada Tabel III.16, pelanggan yang memiliki jarak terkecil adalah **JJ-0008** dan **JJ-0010** dengan perbedaan jarak 0,2. Kedua pelanggan tersebut kemudian digabungkan ke dalam sebuah kelompok baru dengan nama $\mathbf{K_1}$.

3. Langkah ke-3

Kemudian, dilakukan perhitungan jarak antara kelompok yang baru dengan kelompok yang lainnya. Metode yang digunakan adalah metode *complete linkage.* Jarak baru yang dihasilkan merupakan jarak terbesar antar anggota pada dua

kelompok. Persamaan III.1 merupakan formula untuk menghitung jarak dengan metode *complete linkage*.

$$dist(C_i, C_j) = \max_{x_i \in C_i, x_j \in C_j} \{dist(x_i, x_j)\} \quad \text{(III.1)}$$

Berikut merupakan perhitungan jarak antara **K1** dengan beberapa kelompok lainnya.

a. Perhitungan jarak antara **K1** dan **JJ-0001**.

$$dist(JJ\,0001, K_1) = max(dist(JJ\,0001, JJ\,0008), dist(JJ\,0001, JJ\,0010))$$
$$dist(JJ\,0001, K_1) = max(1, 0.857) = 1$$

b. Perhitungan jarak antara **K1** dan **JJ-0005**.

$$dist(JJ\,0005, K_1) = max(dist(JJ\,0005, JJ\,0008), dist(JJ\,0005, JJ\,0010))$$
$$dist(JJ\,0005, K_1) = max(0.714, 0.75) = 0.75$$

c. Perhitungan jarak antara **K1** dan **JJ-0006**.

$$dist(JJ\,0006, K_1) = max(dist(JJ\,0006, JJ\,0008), dist(JJ\,0006, JJ\,0010))$$
$$dist(JJ\,0006, K_1) = max(0.857, 0.714) = 0.857$$

d. Perhitungan jarak antara **K1** dan **JJ-0007**.

$$dist(JJ\,0007, K_1) = max(dist(JJ\,0007, JJ\,0008), dist(JJ\,0007, JJ\,0010))$$
$$dist(JJ\,0007, K_1) = max(0.857, 0.875) = 0.875$$

e. Perhitungan jarak antara **K1** dan **JJ-0009**.

$$dist(JJ\,0009, K_1) = max(dist(JJ\,0009, JJ\,0008), dist(JJ\,0009, JJ\,0010))$$
$$dist(JJ\,0009, K_1) = max(1, 1) = 1$$

f. Perhitungan jarak antara **K1** dan **JJ-0011**.

$$dist(JJ\,0011, K_1) = max(dist(JJ\,0011, JJ\,0008), dist(JJ\,0011, JJ\,0010))$$
$$dist(JJ\,0011, K_1) = max(0.857, 0.875) = 0.875$$

g. Perhitungan jarak antara **K1** dan **JJ-0012**.

$$dist(JJ\,0012, K_1) = max(dist(JJ\,0012, JJ\,0008), dist(JJ\,0012, JJ\,0010))$$
$$dist(JJ\,0012, K_1) = max(1, 1) = 1$$

h. Perhitungan jarak antara **K1** dan **JJ-0015**.

$$dist(JJ\,0015, K_1) = max(dist(JJ\,0015, JJ\,0008), dist(JJ\,0015, JJ\,0010))$$
$$dist(JJ\,0015, K_1) = max(1, 1) = 1$$

i. Perhitungan jarak antara $\mathbf{K_1}$ dan **JJ-0974**.

$$dist(JJ\,0974, K_1) = max(dist(JJ\,0974, JJ\,0008), dist(JJ\,0974, JJ\,0010))$$
$$dist(JJ\,09974, K_1) = max(0.833, 0.857) = 0.857$$

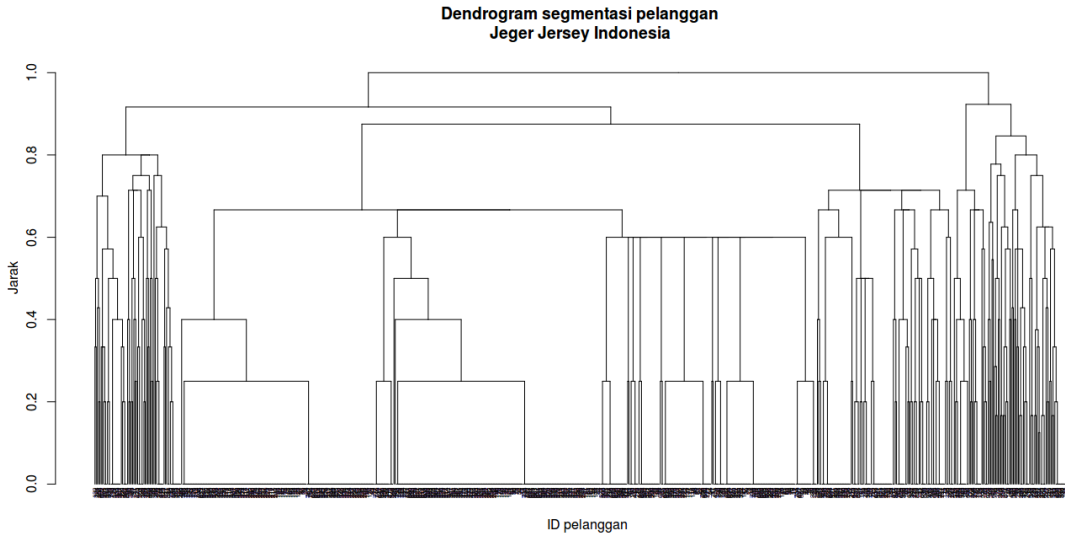
Tabel III.17 merupakan matriks hasil perhitungan jarak menggunakan metode *complete linkage* pada iterasi ke-1.

Tabel III.17 Matrik jarak dengan metode *complete linkage*

	JJ-0001	**JJ-0005**	**JJ-0006**	**JJ-0007**	**JJ-0008/JJ-0010**	**JJ-0009**	**JJ-0011**	**JJ-0012**	**JJ-0014**	. . .	**JJ-0974**
JJ-0001	0										
JJ-0005	1	0									
JJ-0006	0.25	1	0								
JJ-0007	0.6	0.875	0.4	0							
JJ-0008/JJ-0010	1	0.75	0.857	0.875	0						
JJ-0009	0.5	1	0.6	0.6	1	0					
JJ-0011	0.6	1	0.4	0.4	0.875	0.6	0				
JJ-0012	0.5	0.875	0.6	0.6	1	0.5	0.6	0			
JJ-0015	0.5	0.857	0.6	0.25	1	0.5	0.6	0.5	0		
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	
JJ-0974	0.8	0.666	0.833	0.833	0.857	0.8	0.833	0.5	0.8	. . .	0

4. Langkah ke-4

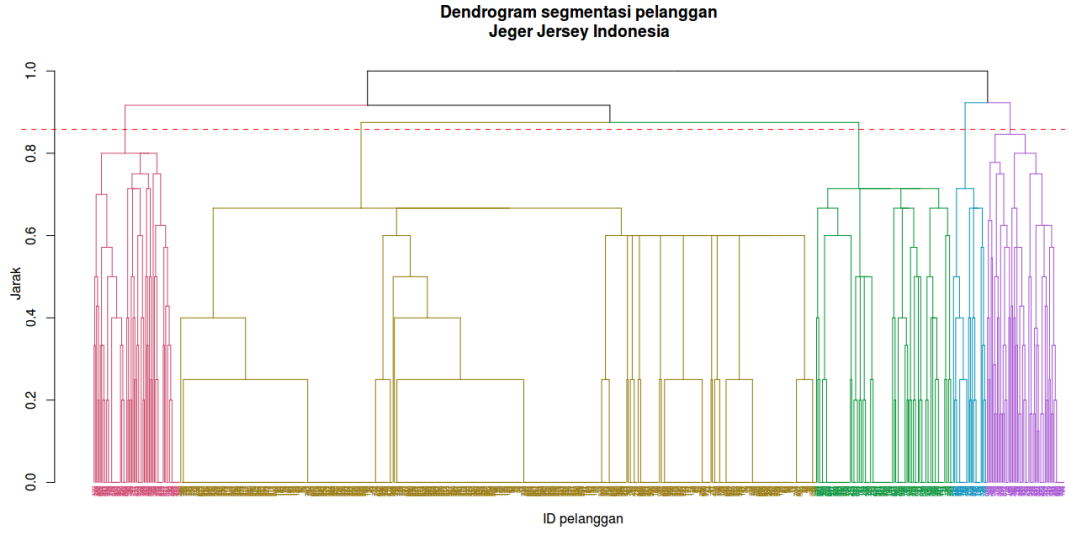
Ulangi langkah kedua dan ketiga sampai tersisa satu kelompok (kumpulan semua pelanggan). Pengulangan akan berhenti pada iterasi ke-772. Gambar III.8 merupakan hasil segmentasi pelanggan yang disajikan dalam sebuah *dendrogram.* Sumbu *x* merepresentasikan pelanggan dan sumbu *y* merepresentasikan jarak antar pelanggan.



Gambar III.8 *Dendrogram* hasil segmentasi pelanggan

5. Langkah ke-5

Langkah terakhir pada metode *hierarchical agglomerative clustering* adalah menentukan jumlah segmen yang diinginkan dengan memotong *dendrogram* pada jarak tertentu. Peneliti mencoba beberapa kemungkinan jumlah segmen yang dapat terbentuk sampai menghasilkan suatu pola atau pengetahuan. Gambar III.9 merupakan *dendrogram* yang dipotong pada jarak 0.86. Pemotongan *dendrogram* tersebut menghasilkan 5 segmen pelanggan.



Gambar III.9 *Dendrogram* yang menghasilkan 5 segmen

Berikut merupakan analisis terhadap segmen pelanggan yang terbentuk dari *dendrogram* pada Gambar III.9.

a. Segmen ke-1

Karakter dari segmen ini adalah pelanggan yang membeli lebih dari tiga produk dalam satu kali transaksi. Jumlah pelanggan yang termasuk ke dalam segmen ke-1 adalah 69 pelanggan. Sebanyak 42% pelanggan pada segmen ini melakukan pembelian *custom name*. Produk yang paling banyak dibeli adalah *jersey* Persib (48% pelanggan), *jersey* Persija (42% pelanggan), dan *jersey* Arema Chronus (19% pelanggan).

b. Segmen ke-2

Karakter dari segmen ini adalah pelanggan yang membeli satu produk dalam satu kali transaksi. Jumlah pelanggan yang termasuk ke dalam segmen ke-2 adalah 506 pelanggan. Sebanyak 34% pelanggan pada segmen ini melakukan pembelian *custom name*. Produk yang paling banyak dibeli adalah *jersey* Persib (32% pelanggan), *jersey* Persija (30% pelanggan), dan *jersey* Persipura (11% pelanggan).

c. Segmen ke-3

Karakter dari segmen ini adalah pelanggan yang membeli dua produk dalam satu kali transaksi. Jumlah pelanggan yang termasuk ke dalam segmen ke-3 adalah 109 pelanggan. Sebanyak 29% pelanggan pada segmen ini melakukan pembelian *custom name*. Produk yang paling banyak dibeli adalah *jersey* Persija (32% pelanggan), *jersey* Persib (31% pelanggan), dan *jersey* Sriwijaya FC (15% pelanggan).

d. Segmen ke-4

Karakter dari segmen ini adalah pelanggan yang membeli dua produk dalam dua kali transaksi. Jumlah pelanggan yang termasuk ke dalam segmen ke-4 adalah 27 pelanggan. Sebanyak 59% pelanggan pada segmen ini melakukan pembelian *custom name*. Produk yang paling banyak dibeli adalah *jersey* Persija (52% pelanggan) dan *jersey* Persib (37% pelanggan).

e. Segmen ke-5

Karakter dari segmen ini adalah pelanggan yang membeli lebih dari tiga produk dalam dua kali transaksi atau lebih. Jumlah pelanggan yang termasuk ke dalam segmen ke-5 sebanyak 62 pelanggan. Sebanyak 77% pelanggan pada segmen ini melakukan pembelian *custom name.* Produk yang paling banyak dibeli adalah *jersey* Persib (65% pelanggan), *jersey* Persija (45% pelanggan), *jersey* Sriwijaya (32% pelanggan), dan *jersey* Persipura (31% pelanggan).

d. Penilaian Model

Berdasarkan analisis terhadap segmen pelanggan yang terbentuk, atribut *jumlah_item* dan *jumlah_transaksi* memiliki pengaruh yang besar terhadap proses segmentasi pelanggan. Pelanggan yang memiliki nilai *jumlah_item* dan *jumlah_transaksi* yang sama akan tergabung pada segmen yang sama. Jika diberikan data pelanggan baru, segmentasi pelanggan akan mudah dilakukan hanya dengan melihat kedua atribut tersebut. Selain itu, setiap segmen yang terbentuk memiliki karakteristik yang berbeda. Oleh karena itu, model ini telah memenuhi kriteria sukses *data mining* yang telah ditentukan. Namun, masih perlu dilakukan evaluasi agar menghasilkan segmentasi pelanggan yang berkualitas.

III.1.6. Spesifikasi Kebutuhan Perangkat Lunak

Spesifikasi kebutuhan perangkat lunak perlu disusun untuk menjelaskan hal-hal yang harus dipenuhi oleh perangkat lunak yang akan dibangun. Spesifikasi kebutuhan perangkat lunak meliputi spesifikasi kebutuhan non-fungsional dan spesifikasi kebutuhan fungsional. Dalam laporan penelitian ini, perangkat lunak yang akan dibangun selanjutnya akan peneliti sebut sebagai sistem segmentasi pelanggan.

III.1.6.1. Spesifikasi Kebutuhan Non Fungsional

Spesifikasi kebutuhan non-fungsional merupakan penentuan kualitas dari perangkat lunak secara keseluruhan. Pada tahap ini, peneliti menentukan batasan-batasan dari fungsi pada sistem segmentasi pelanggan. Spesifikasi kebutuhan non-fungsional dari sistem segmentasi pelanggan disajikan pada Tabel III.18.

Tabel III.18 Spesifikasi kebutuhan non-fungsional

Kode SKPL	Keterangan
SKPL-NF-1	Pengguna dari sistem segmentasi pelanggan adalah manajer *marketing*.
SKPL-NF-2	Sistem segmentasi pelanggan berbasis *desktop*.
SKPL-NF-3	Sistem segmentasi pelanggan memiliki antarmuka yang mudah digunakan.
SKPL-NF-4	Sistem segmentasi pelanggan dapat berjalan dengan baik pada perangkat keras yang tersedia.
SKPL-NF-5	Sistem segmentasi pelanggan dapat memberikan pesan kesalahan saat terjadi kegagalan sistem.
SKPL-NF-6	Sistem segmentasi pelanggan memiliki panduan penggunaan (*user manual*).

III.1.6.2. Spesifikasi Kebutuhan Fungsional

Spesifikasi kebutuhan fungsional merupakan penentuan fitur-fitur atau layanan yang harus tersedia pada sistem segmentasi pelanggan. Fitur-fitur tersebut harus dapat menunjang dalam memenuhi tujuan bisnis yang telah ditentukan. Spesifikasi kebutuhan fungsional pada sistem segmentasi pelanggan dijelaskan pada Tabel III.19.

Tabel III.19 Spesifikasi kebutuhan fungsional

Kode SKPL	Keterangan
SKPL-F-1	Perangkat lunak dapat mengimpor data pelanggan dengan format file *.csv.
SKPL-F-2	Perangkat lunak dapat melakukan seleksi atribut data pelanggan.
SKPL-F-3	Perangkat lunak dapat menghapus baris yang memiliki nilai kosong
SKPL-F-4	Perangkat lunak dapat melakukan proses binning atribut.
SKPL-F-5	Perangkat lunak dapat melakukan proses penurunan atribut.
SKPL-F-6	Perangkat lunak dapat melakukan segmentasi pelanggan.
SKPL-F-7	Perangkat lunak dapat menyimpan *dendrogram* hasil segmentasi pelanggan.
SKPL-F-8	Perangkat lunak dapat menyimpan data hasil segmentasi pelanggan.

III.1.7. Analisis Kebutuhan Non Fungsional

Analisis kebutuhan non-fungsional merupakan proses identifikasi dan penentuan perangkat pendukung yang dibutuhkan. Hal tersebut dilakukan agar sistem segmentasi pelanggan dapat diimplementasikan dengan baik. Analisis kebutuhan non-fungsional meliputi analisis kebutuhan pengguna, analisis kebutuhan perangkat keras, dan analisis kebutuhan perangkat lunak.

III.1.7.1. Analisis Kebutuhan Pengguna

Pada tahap ini dilakukan identifikasi terhadap karakteristik pengguna sistem segmentasi pelanggan. Pengguna harus memenuhi beberapa kriteria agar dapat mengoperasikan perangkat lunak dengan baik. Tabel III.20 menyajikan analisis terhadap pengguna yang akan menggunakan perangkat lunak.

Tabel III.20 Analisis kebutuhan pengguna

Entitas	Keterangan
Pengguna	Manajer *marketing* Jeger Jersey Indonesia.
Tanggung jawab	Melakukan analisis pasar, membuat desain promosi, dan memasarkan produk kepada pelanggan.
Hak akses	Akses penuh terhadap data pelanggan dan melakukan segmentasi pelanggan.
Tingkat pendidikan	Sarjana di jurusan Teknik Informatika.
Tingkat keterampilan	Memahami data pelanggan secara menyeluruh dan menguasai statistik data pelanggan.
Pengalaman	Mampu mengoperasikan program komputer dengan baik.

Berdasarkan data pada Tabel III.20, dapat disimpulkan bahwa manajer *marketing* akan dapat mengoperasikan sistem segmentasi pelanggan. Namun, manajer *marketing* tetap dianjurkan untuk membaca dokumentasi penggunaan perangkat lunak. Hal tersebut dilakukan agar manajer *marketing* dapat menggunakan sistem segmentasi pelanggan dengan baik dan benar.

III.1.7.2. Analisis Kebutuhan Perangkat Keras

Analisis terhadap kebutuhan perangkat keras perlu dilakukan agar implementasi sistem segmentasi pelanggan dapat bekerja dengan maksimal. Tabel III.21 menyajikan data perangkat keras yang tersedia saat ini.

Tabel III.21 Analisis kebutuhan perangkat keras

Jenis perangkat keras	Spesifikasi
Prosesor	4 *core* dengan kecepatan 1800 MHz
RAM	4 GB
Harddisk	320 GB
Monitor	14 inci dengan resolusi 1366 x 768 piksel

Berdasarkan data pada Tabel III.21, dapat disimpulkan bahwa perangkat keras yang digunakan saat ini telah memenuhi kebutuhan yang diperlukan untuk menjalankan sistem segmentasi pelanggan. Oleh karena itu, manajer *marketing* tidak perlu melakukan *upgrade* dan/ atau menambahkan perangkat keras lainnya.

III.1.7.3. Analisis Kebutuhan Perangkat Lunak

Pada tahap ini dilakukan analisis terhadap perangkat lunak yang dibutuhkan. Perangkat lunak tersebut harus dipenuhi agar implementasi sistem segmentasi pelanggan dapat berjalan dengan baik. Tabel III.22 menyajikan data perangkat lunak yang tersedia saat ini.

Tabel III.22 Analisis kebutuhan perangkat lunak

Jenis perangkat lunak	Spesifikasi
Sistem operasi	Ubuntu 14.04.2 LTS 64 bit
Kernel	Linux 3.13.0-24
Lingkungan desktop	Unity

Berdasarkan data pada Tabel III.22, perangkat lunak yang dibutuhkan untuk mendukung sistem yang akan dibangun belum terpenuhi. Tabel III.23 menyajikan beberapa perangkat lunak tambahan yang perlu dipasang.

Tabel III.23 Perangkat lunak yang harus dipasang

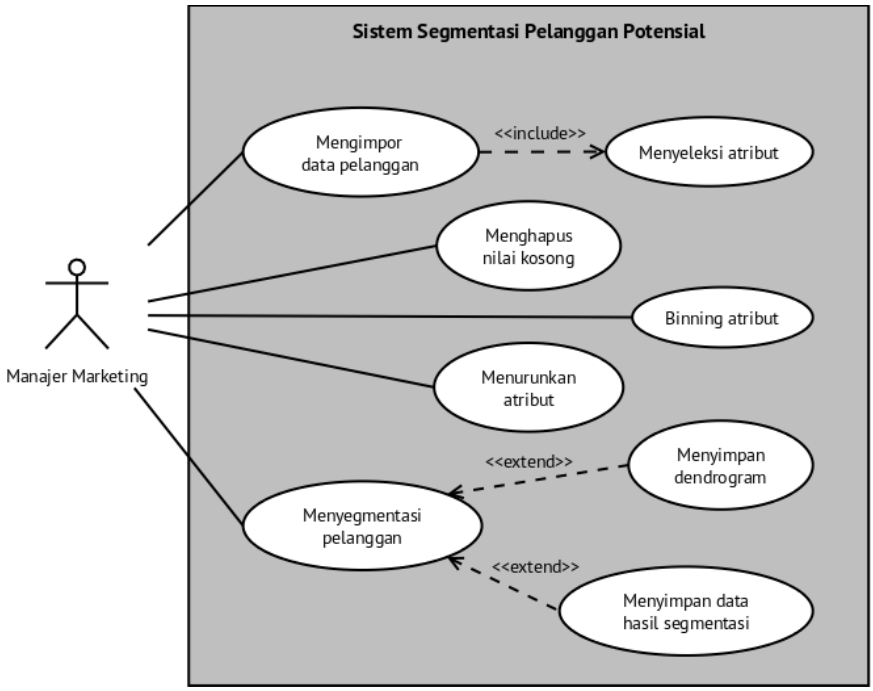
Jenis perangkat lunak	Spesifikasi
Bahasa pemrograman	Python 2.7.8
Library	PyQt 4, Pandas, Scipy, Scikit-Learn, Matplotlib

III.1.8. Analisis Kebutuhan Fungsional

Pada tahap ini, dilakukan pemodelan sistem segmentasi pelanggan. Pemodelan dilakukan terhadap setiap spesifikasi kebutuhan fungsional yang telah ditentukan. Pendekatan analisis yang digunakan adalah pendekatan berbasis objek menggunakan diagram UML 2.0. Diagram yang digunakan meliputi *use case diagram, activity diagram, sequence diagram,* dan *class diagram.*

III.1.8.1. Use Case Diagram

Use case diagram digunakan untuk menggambarkan kebutuhan fungsional dari sistem yang akan dibangun. *Use case diagram* menjelaskan hubungan antara pengguna dengan sistem. Diagram ini menjadi titik awal bagi pengembangan sistem menggunakan pendekatan berorientasi objek. Gambar III.10 merupakan *use case diagram* sistem segmentasi pelanggan.



Gambar III.10 Use case diagram sistem segmentasi pelanggan

Use case diagram tidak memberikan detail yang cukup untuk menjelaskan bagaimana kebutuhan sistem akan terpenuhi. Untuk itu, *use case scenario* digunakan untuk menunjukkan kondisi yang terjadi dan langkah-langkah apa saja yang terlibat pada setiap *use case.*

1. *Use Case Scenario* Mengimpor Data Pelanggan

Use case mengimpor data pelanggan merupakan fungsional untuk mengimpor data dari media penyimpanan ke dalam sistem segmentasi pelanggan. Data tersebut selanjutnya akan diproses untuk menghasilkan segmentasi pelanggan. *Use case scenario* mengimpor data pelanggan dapat dilihat pada Tabel III.24.

Use Case Name	**Mengimpor Data Pelanggan**	
Related Requirements	SKPL-F-01	
Goal In Context	Mengimpor data pelanggan ke dalam *database.*	
Precondition	Aktor berada di halaman utama sistem.	
Successful End Condition	Data pelanggan berhasil diimpor ke dalam *database.*	
Failed End Condition	Data pelanggan gagal diimpor ke dalam *database.*	
Actors	Manajer *Marketing.*	
Trigger	Aktor menekan tombol untuk mengimpor data pelanggan.	
Main Flow	*Step*	*Action*
	1	Aktor menekan tombol untuk mengimpor data pelanggan.
	2	Aktor memilih data pelanggan yang disimpan pada komputer (format data yang diizinkan berekstensi *.csv)
	3	Sistem memverifikasi format data.
	4	Sistem menampilkan atribut-atribut data pelanggan.
	5 **include::Menyeleksi Atribut**	Aktor menyeleksi atribut-atribut yang diperlukan.
	6	Sistem menampilkan data pelanggan yang berhasil diimpor.
Extension	*Step*	*Branching Action*
	3.1	Format data yang diimpor tidak diizinkan.
	3.2	Sistem menolak impor data.

Tabel III.24 *Use case scenario* mengimpor data pelanggan

2. *Use Case Scenario* Menyeleksi Atribut

Use case menyeleksi atribut merupakan fungisonal untuk menyeleksi atribut-atribut pada data yang telah diimpor. Atribut-atribut yang dipilih merupakan data yang akan digunakan pada proses segmentasi. *Use case scenario* menyeleksi atribut dapat dilihat pada Tabel III.25.

Use Case Name	**Menyeleksi Atribut**
Related Requirements	SKPL-F-02
Goal In Context	Menghasilkan data pelanggan dengan atribut-atribut tertentu.
Precondition	Sistem menampilkan atribut-atribut data pelanggan ketika mengimpor data pelanggan.
Successful End Condition	Data pelanggan dengan atribut-atribut yang telah diseleksi berhasil diimpor ke dalam *database.*
Failed End Condition	Data pelanggan dengan atribut-atribut yang telah diseleksi gagal diimpor ke dalam *database.*
Actors	Manajer *Marketing.*
Trigger	Aktor memilih atribut dengan menekan *checkbox* pada atribut-atribut yang akan digunakan pada proses segmentasi pelanggan.

Main Flow	*Step*	*Action*
	1	Aktor memilih atribut dengan menekan *checkbox* pada atribut-atribut yang akan digunakan.
	2	Aktor menekan tombol untuk mengimpor data pelanggan dengan atribut-atribut yang telah diseleksi.
	3	Sistem mengimpor data pelanggan dengan atribut-atribut yang telah diseleksi ke dalam *database*.
Extension	*Step*	*Branching Action*
	2.1	Aktor tidak memilih atribut.
	2.2	Sistem menolak impor data pelanggan.

Tabel III.25 *Use case scenario* menyeleksi atribut

3. *Use Case Scenario* Menghapus Nilai Kosong

Use case menghapus nilai kosong merupakan fungsional untuk menghapus nilai kosong yang terdapat pada data pelanggan. *Use case scenario* menghapus nilai kosong dapat dilihat pada Tabel III.26.

Use Case Name	**Menghapus Nilai Kosong**	
Related Requirements	SKPL-F-03	
Goal In Context	Melakukan penghapusan nilai kosong pada data pelanggan.	
Precondition	Data pelanggan sudah diimpor.	
Successful End Condition	Data pelanggan yang memiliki nilai kosong berhasil dihapus.	
Failed End Condition	Data pelanggan yang memiliki nilai kosong gagal dihapus.	
Actors	Manajer *Marketing*	
Trigger	Aktor menekan tombol untuk menghapus nilai kosong.	
Main Flow	*Step*	*Action*
	1	Aktor menekan tombol untuk menghapus nilai kosong
	2	Sistem melakukan pengecekan data yang memiliki nilai kosong.
	3	Sistem menghapus data yang memiliki nilai kosong.
Extension	*Step*	*Branching Action*
	2.1	Tidak ditemukan data yang memiliki nilai kosong.
	2.2	Sistem membatalkan penghapusan data.

Tabel III.26 *Use case scenario* menghapus nilai kosong

4. *Use Case Scenario* Binning Atribut

Use case binning atribut merupakan fungsional untuk melakukan *binning* atribut pada data pelanggan yang telah diimpor ke dalam sistem. *Use case scenario* binning atribut dapat dilihat pada Tabel III.27.

Use Case Name	**Binning Atribut**
Related Requirements	SKPL-F-04
Goal In Context	Melakukan *binning* atribut pada data pelanggan.
Precondition	Data pelanggan tidak memiliki nilai kosong.
Successful End Condition	Binning atribut berhasil dilakukan.
Failed End Condition	Binning atribut gagal dilakukan.
Actors	Manajer *Marketing*
Trigger	Aktor menekan tombol untuk melakukan *binning* atribut.

Main Flow	*Step*	*Action*
	1	Aktor menekan tombol untuk melakukan *binning* atribut.
	2	Aktor memilih atribut yang akan dilakukan *binning*.
	3	Aktor menentukan jumlah *bin* yang diinginkan.
	4	Sistem melakukan *binning* atribut
Extension	*Step*	*Branching Action*
	4.1	Jumlah bin tidak diizinkan
	4.2	Sistem gagal melakukan binning atribut.

Tabel III.27 *Use case scenario* binning atribut

5. *Use Case Scenario* Menurunkan Atribut

Use case menurunkan atribut merupakan fungsional untuk menurunkan atribut pada data pelanggan yang telah diimpor ke dalam sistem. *Use case scenario* menurunkan atribut dapat dilihat pada Tabel III.28.

Use Case Name	**Binning Atribut**	
Related Requirements	SKPL-F-05	
Goal In Context	Melakukan penurunan atribut pada data pelanggan.	
Precondition	Data pelanggan tidak memiliki nilai kosong.	
Successful End Condition	Penurunan atribut berhasil dilakukan.	
Failed End Condition	Penurunan atribut gagal dilakukan.	
Actors	Manajer *Marketing*	
Trigger	Aktor menekan tombol untuk melakukan penurunan atribut.	
Main Flow	*Step*	*Action*
	1	Aktor menekan tombol untuk melakukan penurunan atribut.
	2	Aktor memilih atribut yang akan diturunkan.
	3	Sistem melakukan penurunan atribut.
Extension	*Step*	*Branching Action*
	3.1	Sistem gagal melakukan penurunan atribut.

Tabel III.28 *Use case scenario* binning atribut

6. *Use Case Scenario* Menyegmentasi Pelanggan

Use case menyegmentasi pelanggan merupakan fungsional untuk melakukan segmentasi menggunakan metode *hierarchical agglomerative clustering*. *Use case scenario* menyegmentasi pelanggan dapat dilihat pada Tabel III.29.

Use Case Name	**Menyegmentasi Pelanggan**
Related Requirements	SKPL-F-06
Goal In Context	Melakukan segmentasi pelanggan menggunakan metode *hierarchical agglomerative clustering*.
Precondition	Data pelanggan sudah dilakukan pembersihan.
Successful End Condition	Menghasilkan segmen-segmen pelanggan.
Failed End Condition	Data pelanggan gagal disegmentasikan.
Actors	Manajer *Marketing*
Trigger	Aktor menekan tombol untuk menyegmentasikan data pelanggan yang sudah dibersihkan.

Main Flow	*Step*	*Action*
	1	Aktor menekan tombol untuk menyegmentasikan data pelanggan yang sudah dibersihkan.
	2	Sistem melakukan verifikasi kecocokan data dengan metode yang digunakan.
	3	Sistem melakukan segmentasi pelanggan menggunakan metode *hierarchical agglomerative clustering.*
	4	Sistem menampilkan *dendrogram* hasil segmentasi pelanggan.
	5	Sistem menampilkan data hasil segmentasi pelanggan.
	6	Aktor menyimpan *dendrogram* hasil segmentasi pelanggan.
	7	Aktor menyimpan data hasil segmentasi pelanggan.
Extension	*Step*	*Branching Action*
	2.1	Data tidak cocok dengan metode yang digunakan.
	2.2	Sistem membatalkan segmentasi pelanggan.

Tabel III.29 *Use case scenario* menyegmentasi pelanggan

7. *Use Case Scenario* Menyimpan Dendrogram

Use case menyimpan dendrogram merupakan fungsional untuk menyimpan *dendrogram* hasil segmentasi pelanggan. *Dendrogram* merupakan visualisasi hasil segmentasi pelanggan. *Use case scenario* menyimpan dendrogram dapat dilihat pada Tabel III.30.

Use Case Name	**Menyimpan Dendrogram**	
Related Requirements	SKPL-F-07	
Goal In Context	Menyimpan *dendrogram* hasil segmentasi pelanggan.	
Precondition	*Dendrogram* sudah ditampilkan.	
Successful End Condition	*Dendrogram* berhasil disimpan.	
Failed End Condition	*Dendrogram* gagal disimpan.	
Actors	Manajer *Marketing*	
Trigger	Aktor menekan tombol untuk menyimpan *dendrogram.*	
Main Flow	*Step*	*Action*
	1	Aktor menekan tombol untuk menyimpan *dendrogram.*
	2	Aktor menentukan lokasi pada *disk* untuk menyimpan *dendrogram.*
	3	Aktor memilih format *file* untuk *dendrogram.*
	4	Aktor menekan tombol simpan.
	5	Sistem memverifikasi format *file* yang diizinkan (*.png dan *.pdf)
	6	Sistem menyimpan *dendrogram* pada komputer.
Extension	*Step*	*Branching Action*
	5.1	Format *file* tidak diizinkan.
	5.2	Sistem membatalkan penyimpanan *dendrogram.*

Tabel III.30 *Use case scenario* menyimpan dendrogram

8. *Use Case Scenario* Menyimpan Data Hasil Segmentasi

Use case ini merupakan fungsional untuk menyimpan data hasil segmentasi pelanggan. *Use case scenario* tersebut dapat dilihat pada Tabel III.31.

Use Case Name	**Menyimpan Data Hasil Segmentasi**	
Related Requirements	SKPL-F-08	
Goal In Context	Menyimpan Data Hasil Segmentasi pelanggan.	
Precondition	Data pelanggan sudah ditampilkan.	
Successful End Condition	Data Hasil Segmentasi pelanggan berhasil disimpan.	
Failed End Condition	Data Hasil Segmentasi pelanggan gagal disimpan.	
Actors	Manajer *Marketing*	
Trigger	Aktor menekan tombol untuk menyimpan data hasil segmentasi.	
Main Flow	*Step*	*Action*
	1	Aktor menekan tombol untuk menyimpan data hasil segmentasi.
	2	Aktor menentukan lokasi pada *disk* untuk menyimpan data hasil segmentasi.
	3	Aktor memilih format *file* untuk data hasil segmentasi.
	4	Aktor menekan tombol simpan.
	5	Sistem memverifikasi format *file* yang diizinkan (*.csv)
	6	Sistem menyimpan data hasil segmentasi.
Extension	*Step*	*Branching Action*
	5.1	Format *file* tidak diizinkan.
	5.2	Sistem membatalkan penyimpanan data hasil segmentasi.

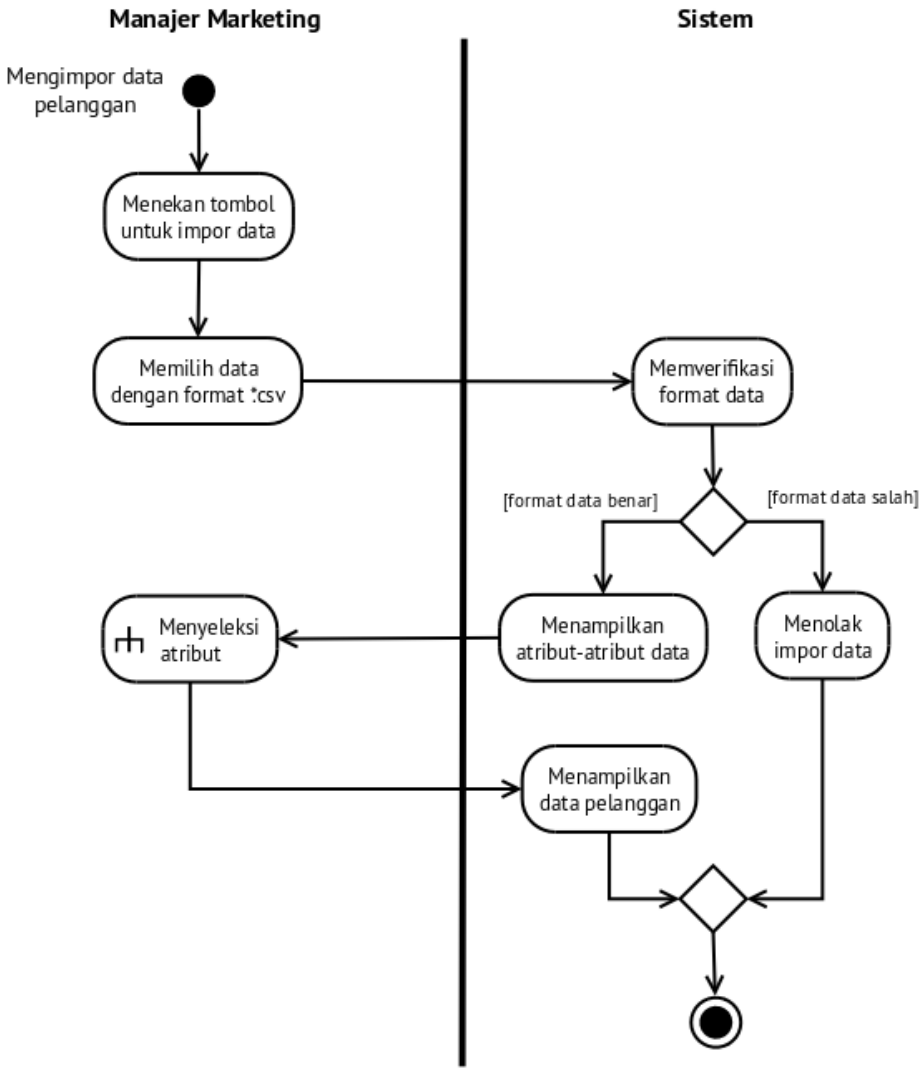
Tabel III.31 *Use case scenario* menyimpan data hasil segmentasi

III.1.8.2. Activity Diagram

Activity diagram digunakan untuk menggambarkan langkah-langkah pada setiap kebutuhan yang telah ditentukan pada *use case diagram.* Diagram ini menggambarkan setiap proses yang terjadi pada sistem, baik proses yang sekuensial atau proses yang dilakukan secara paralel. Berikut merupakan *activity diagram* untuk setiap *use case* yang telah ditetapkan.

1. *Activity Diagram* Mengimpor Data Pelanggan

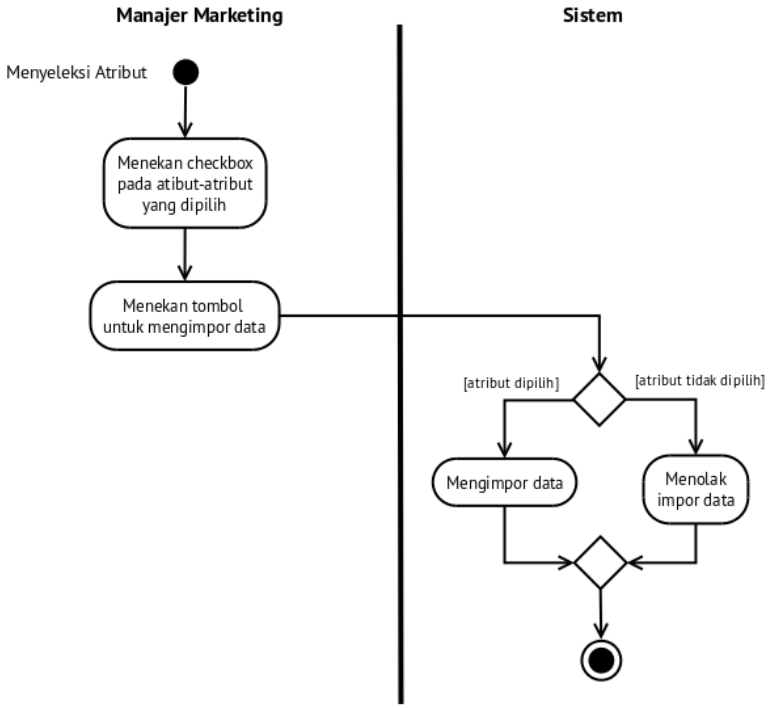
Activity Diagram mengimpor data pelanggan menggambarkan aktivitas aktor dan sistem dalam proses impor data pelanggan. Proses impor data diawali dengan pengguna yang memilih data pelanggan pada media penyimpanan dan diakhiri dengan sistem yang menampilkan data tersebut. *Activity Diagram* mengimpor data pelanggan dapat dilihat pada Gambar III.11.



Gambar III.11 *Activity diagram* mengimpor data pelanggan

2. *Activity Diagram* Menyeleksi Atribut

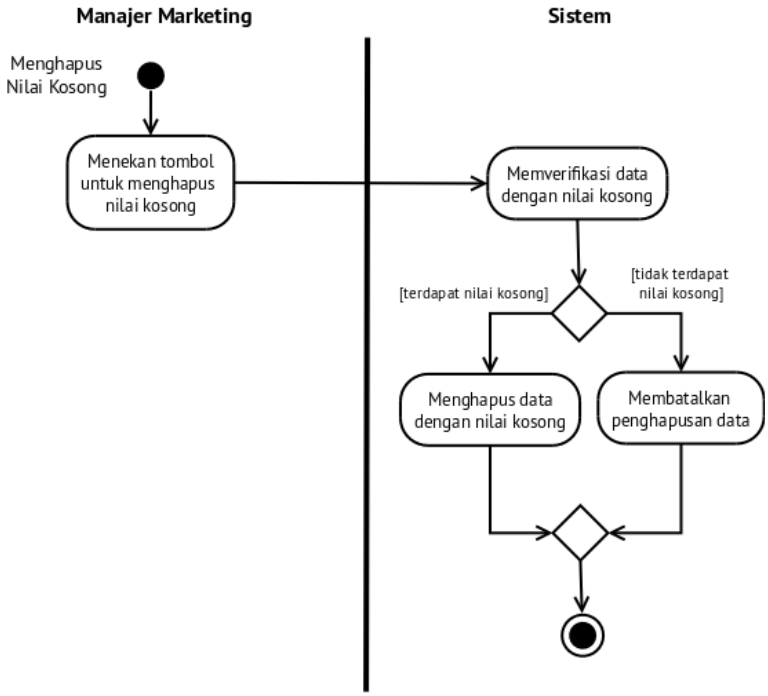
Activity Diagram menyeleksi atribut menggambarkan aktivitas aktor dan sistem dalam proses seleksi atribut data pelanggan. Aktivitas menyeleksi atribut merupakan bagian dari proses mengimpor data pelanggan. Aktivitas ini diawali dengan pengguna yang menekan *checkbox* pada atribut-atribut yang akan digunakan dan diakhiri dengan pengimporan data oleh sistem. *Activity Diagram* menyeleksi atribut dapat dilihat pada Gambar III.12.



Gambar III.12 *Activity diagram* menyeleksi atribut

3. *Activity Diagram* Menghapus Nilai Kosong

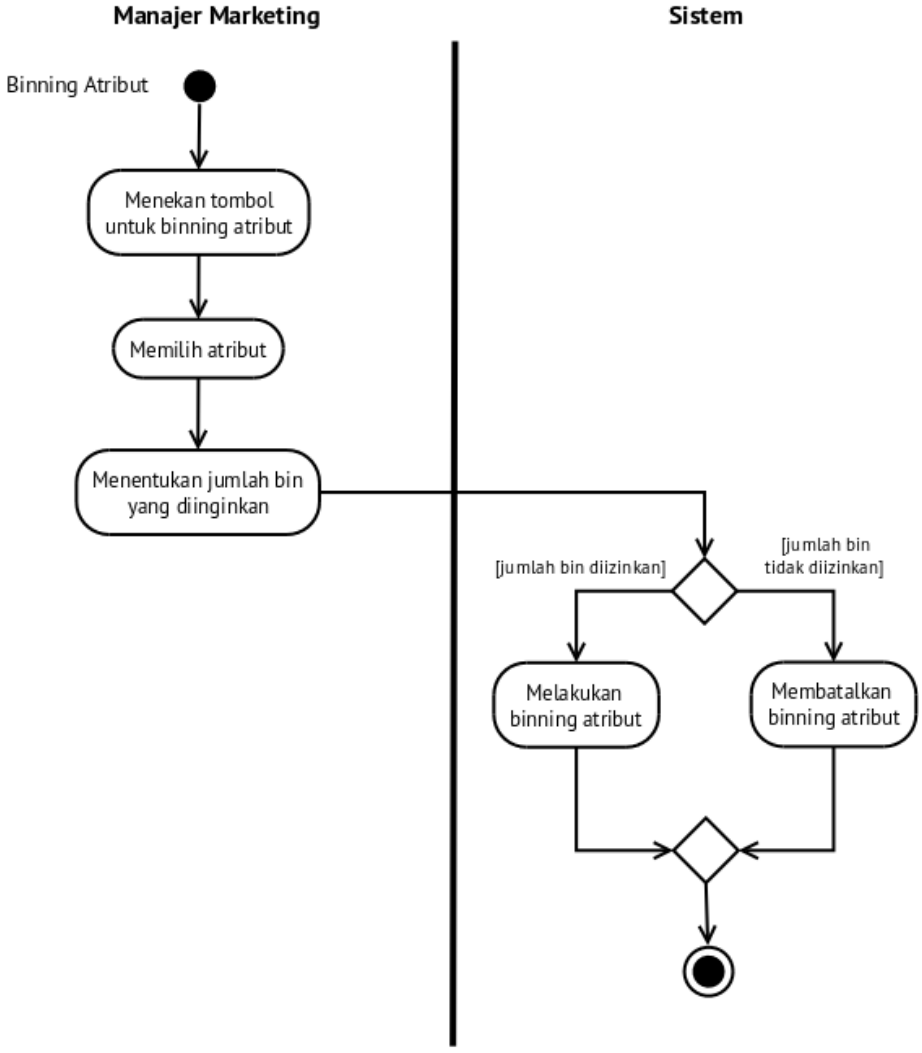
Activity Diagram ini menggambarkan prosedur penghapusan nilai kosong. *Activity Diagram* menghapus nilai kosong dapat dilihat pada Gambar III.13.



Gambar III.13 *Activity diagram* menghapus nilai kosong

4. *Activity Diagram* Binning Atribut

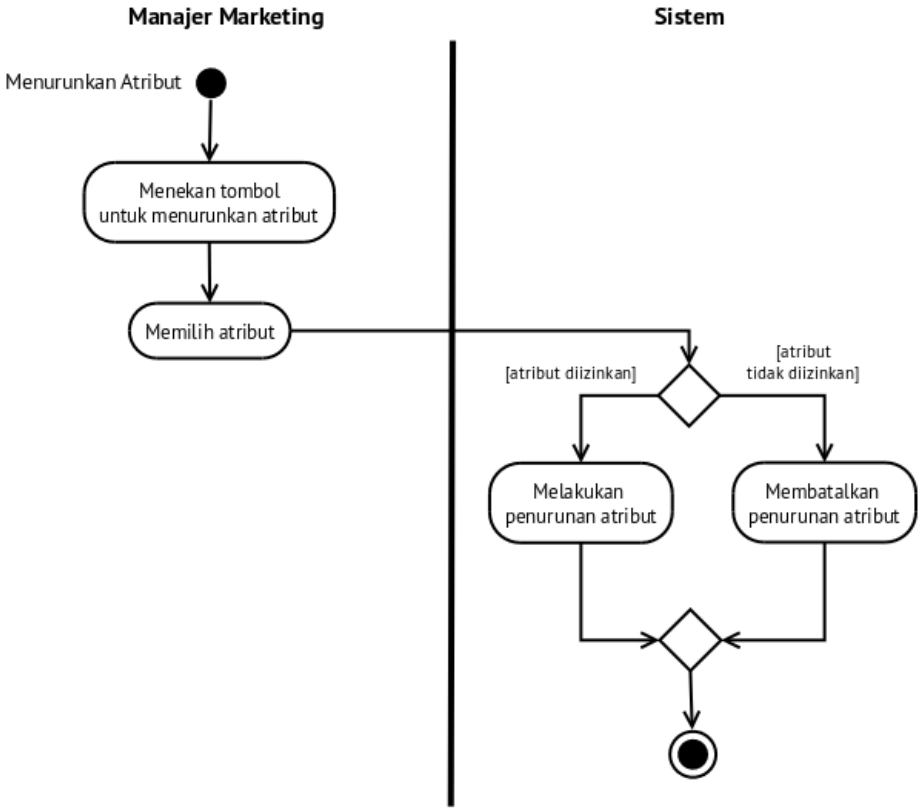
Activity Diagram binning atribut menggambarkan aktivitas aktor dan sistem dalam proses binning atribut. Aktivitas ini diawali dengan pengguna yang menekan tombol untuk *binning* atribut, memilih atribut yang akan dilakukan *binning,* dan menentukan jumlah *bin.* Jika jumlah *bin* diizinkan oleh sistem, maka proses *binning* akan dilakukan. Namun jika jumlah *bin* tidak diizinkan oleh sistem maka proses *binning* akan dibatalkan. *Activity Diagram* binning atribut dapat dilihat pada Gambar III.14.



Gambar III.14 *Activity diagram* binning atribut

5. *Activity Diagram* Menurunkan Atribut

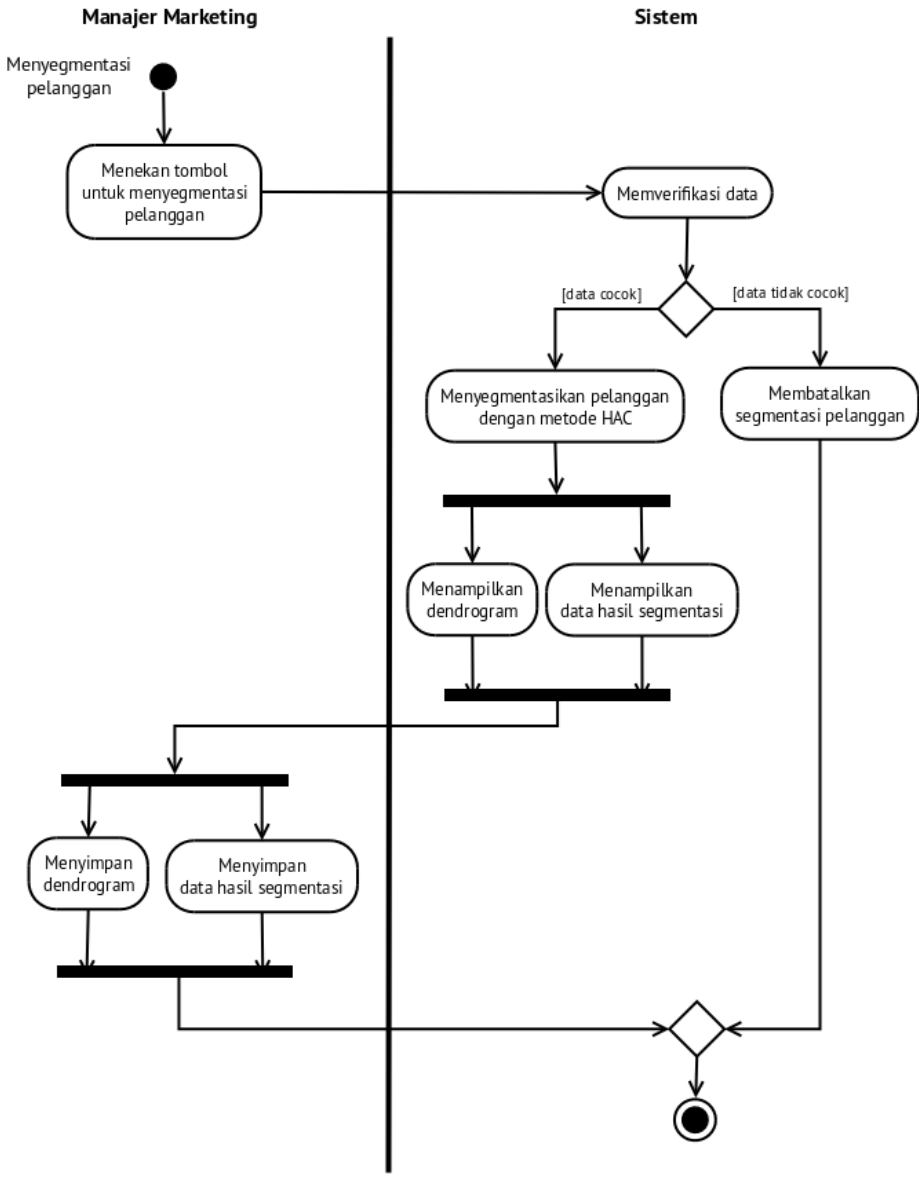
Activity Diagram menurunkan atribut menggambarkan aktivitas aktor dan sistem dalam proses menurunkan atribut yang sudah ada menjadi atribut-atribut baru. Aktivitas ini diawali dengan pengguna yang menekan tombol untuk menurunkan atribut dan memilih atribut yang akan diturunkan. Jika atribut yang dipilih diizinkan oleh sistem, maka proses penurunan atribut akan dilakukan. Namun jika atribut yang dipilih tidak diizinkan oleh sistem, maka proses penurunan atribut akan dibatalkan. *Activity Diagram* menurunkan atribut disajikan pada Gambar III.15.



Gambar III.15 *Activity diagram* menurunkan atribut

6. *Activity Diagram* Menyegmentasi Pelanggan

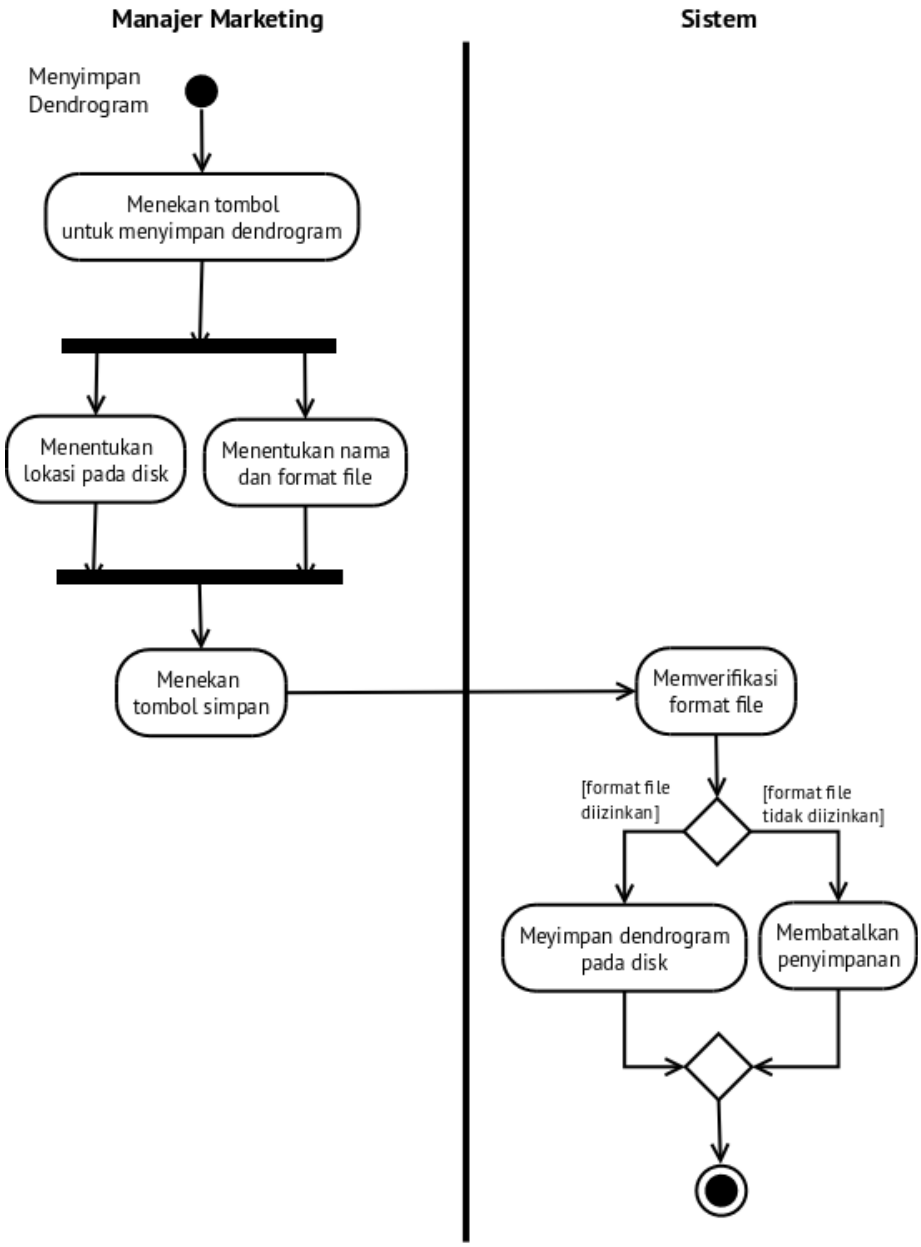
Activity Diagram menyegmentasi pelanggan menggambarkan aktivitas aktor dan sistem dalam proses segmentasi pelanggan dan menampilkan hasil segmentasinya. Pada aktivitas ini akan diterapkan metode *hierarchical agglomerative clustering*. *Activity Diagram* segmentasi pelanggan dapat dilihat pada Gambar III.16.



Gambar III.16 *Activity diagram* menyegmentasi pelanggan

7. *Activity Diagram* Menyimpan Dendrogram

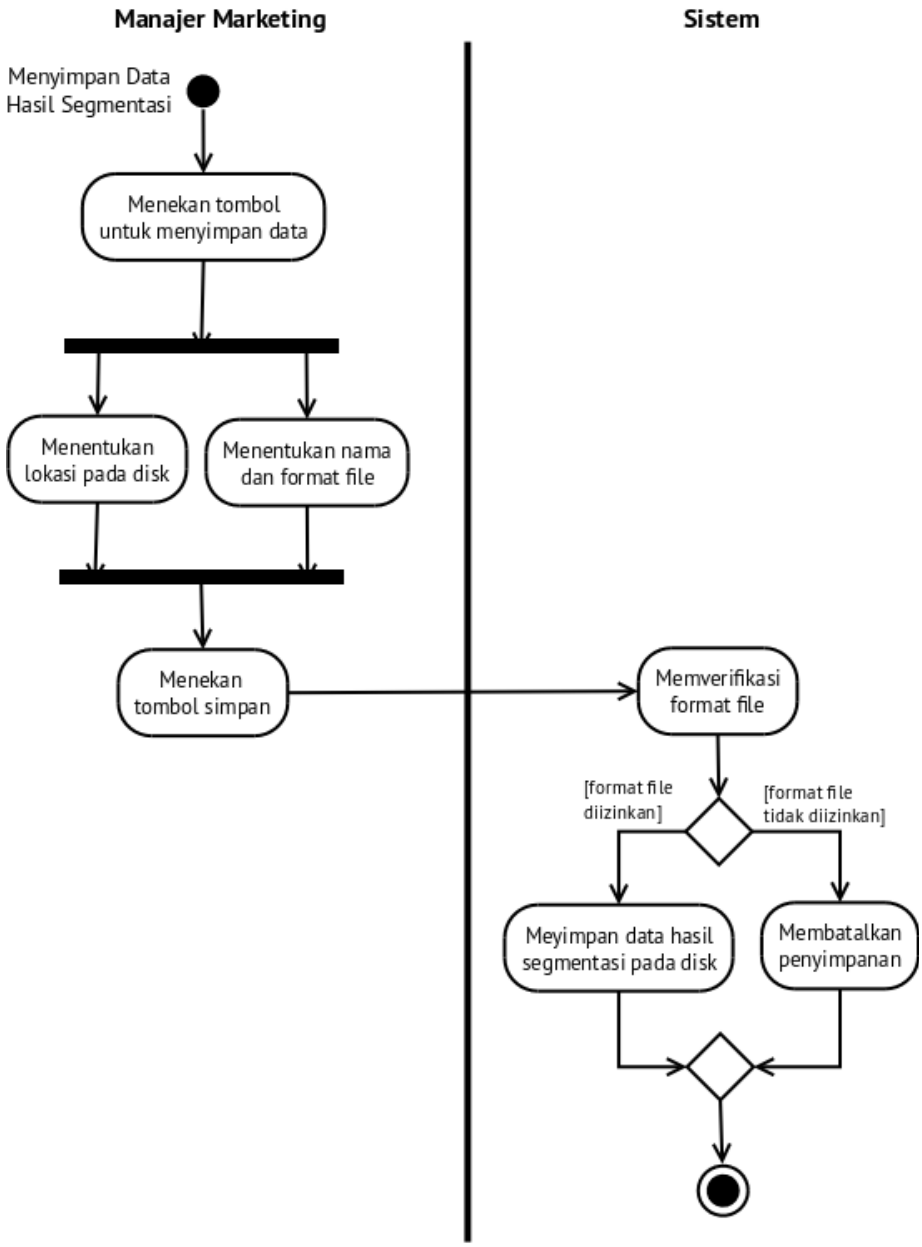
Activity Diagram menyimpan dendrogram menggambarkan aktivitas aktor dan sistem dalam proses penyimpanan *dendrogram* hasil segmentasi pelanggan. Format file *dendrogram* yang diizinkan berekstensi *.png. *Activity Diagram* menyimpan dendrogram dapat dilihat pada Gambar III.17.



Gambar III.17: *Activity diagram* menyimpan dendrogram

8. *Activity Diagram* Menyimpan Data Hasil Segmentasi

Activity Diagram menyimpan data hasil segmentasi menggambarkan aktivitas aktor dan sistem dalam proses penyimpanan data hasil segmentasi pelanggan. *Activity Diagram* menyimpan data hasil segmentasi dapat dilihat pada Gambar III.18.



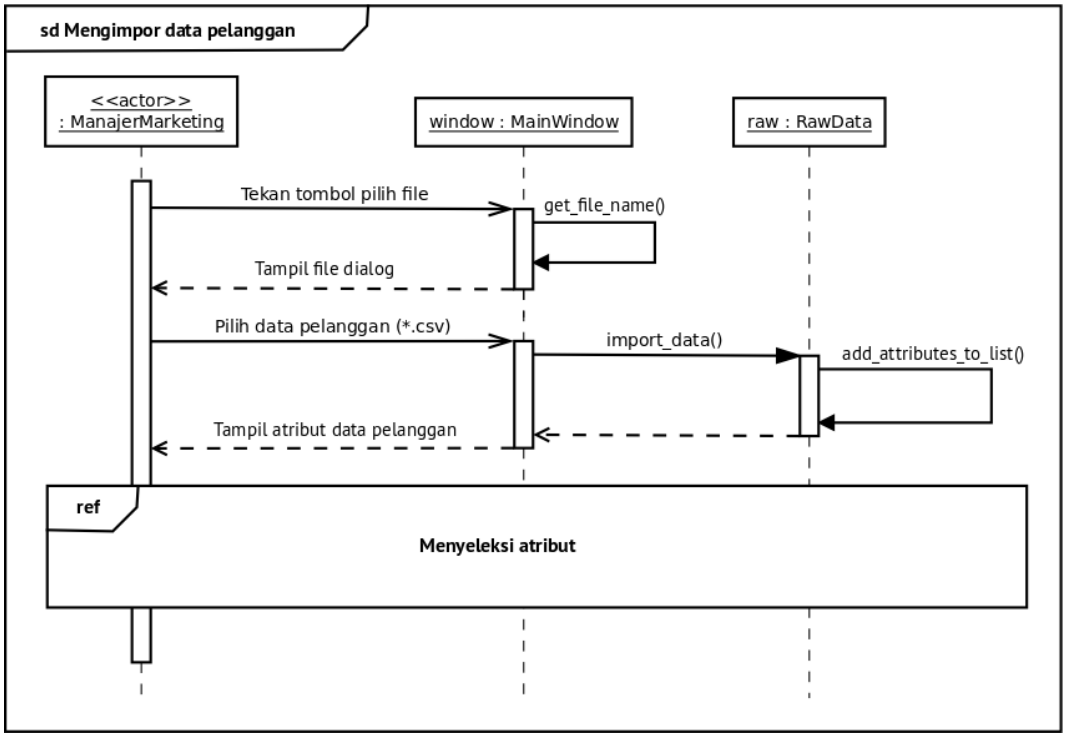
Gambar III.18 *Activity diagram* menyimpan data hasil segmentasi

III.1.8.3. Sequence Diagram

Sequence diagram digunakan untuk menggambarkan urutan interaksi antara objek-objek pada sistem yang akan dibangun. Diagram ini dapat menggambarkan interaksi apa saja yang akan dipicu oleh sebuah *use case* yang dijalankan. Berikut merupakan *sequence diagram* pada sistem segmentasi pelanggan.

1. *Sequence Diagram* Mengimpor Data Pelanggan

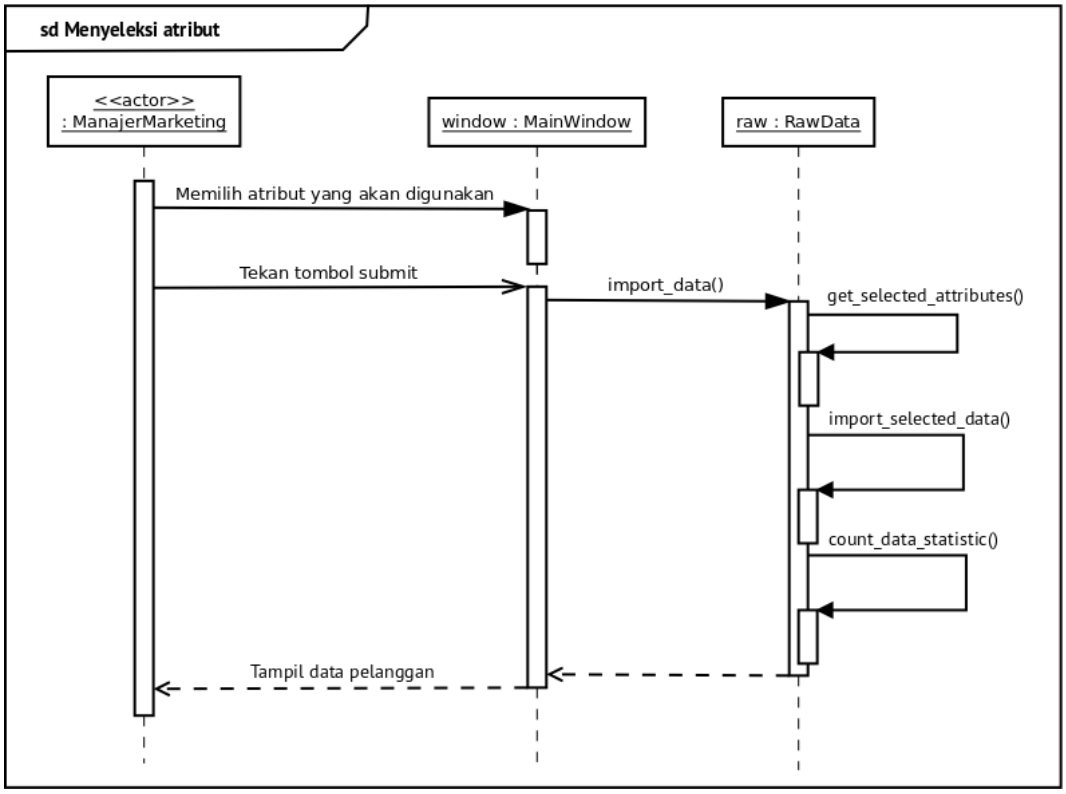
Sequence Diagram mengimpor data pelanggan menggambarkan urutan interaksi antar objek pada proses impor data pelanggan. *Sequence Diagram* mengimpor data pelanggan dapat dilihat pada Gambar III.19.



Gambar III.19 *Sequence diagram* mengimpor data pelanggan

2. *Sequence Diagram* Menyeleksi Atribut

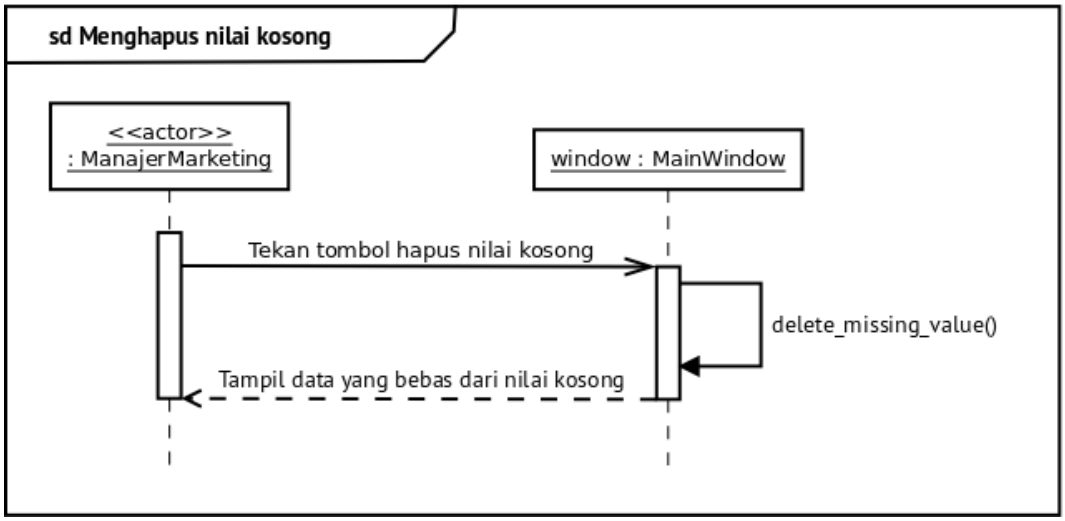
Sequence Diagram menyeleksi atribut menggambarkan urutan interaksi antar objek pada proses pemilihan atribut data pelanggan. *Sequence Diagram* menyeleksi atribut dapat dilihat pada Gambar III.20.



Gambar III.20 *Sequence diagram* menyeleksi atribut

3. *Sequence Diagram* Menghapus Nilai Kosong

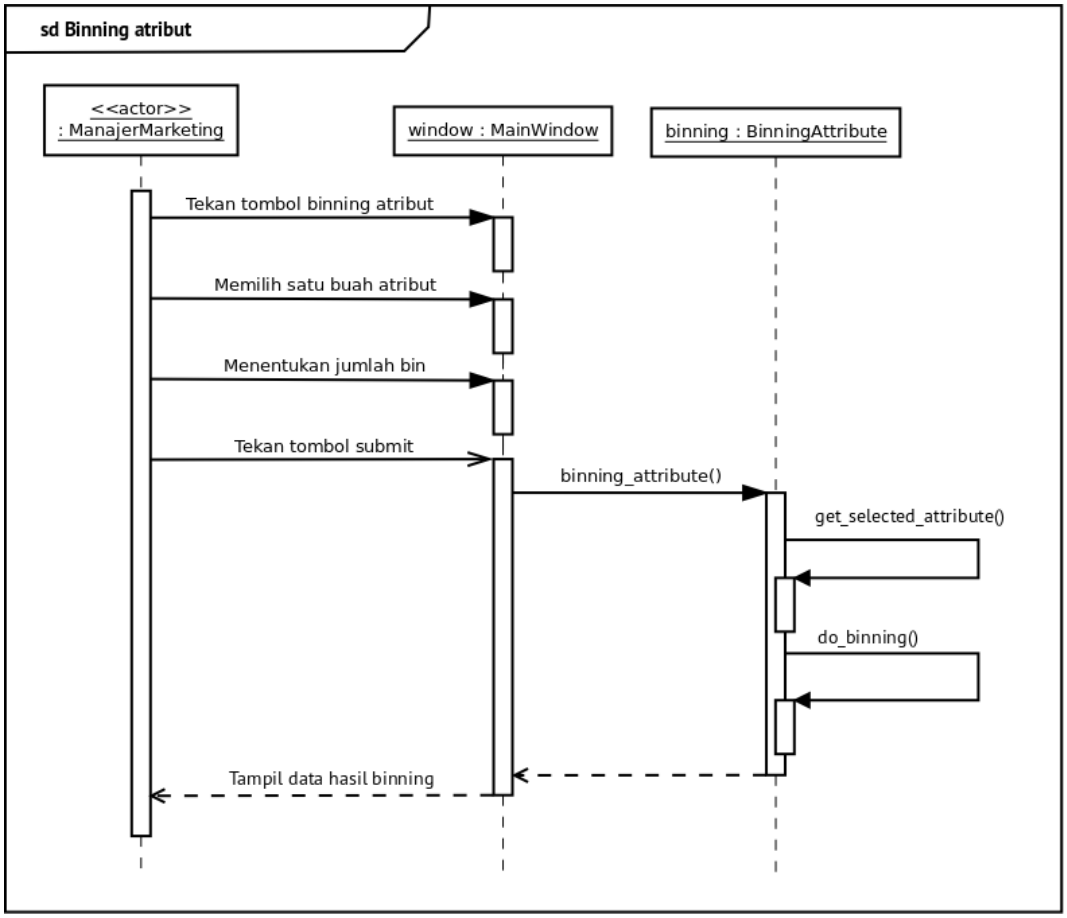
Sequence Diagram ini menggambarkan urutan interaksi antar objek dalam proses penghapusan nilai kosong pada data pelanggan. *Sequence Diagram* menghapus nilai kosong dapat dilihat pada Gambar III.21.



Gambar III.21 *Sequence diagram* menghapus nilai kosong

4. *Sequence Diagram* Binning Atribut

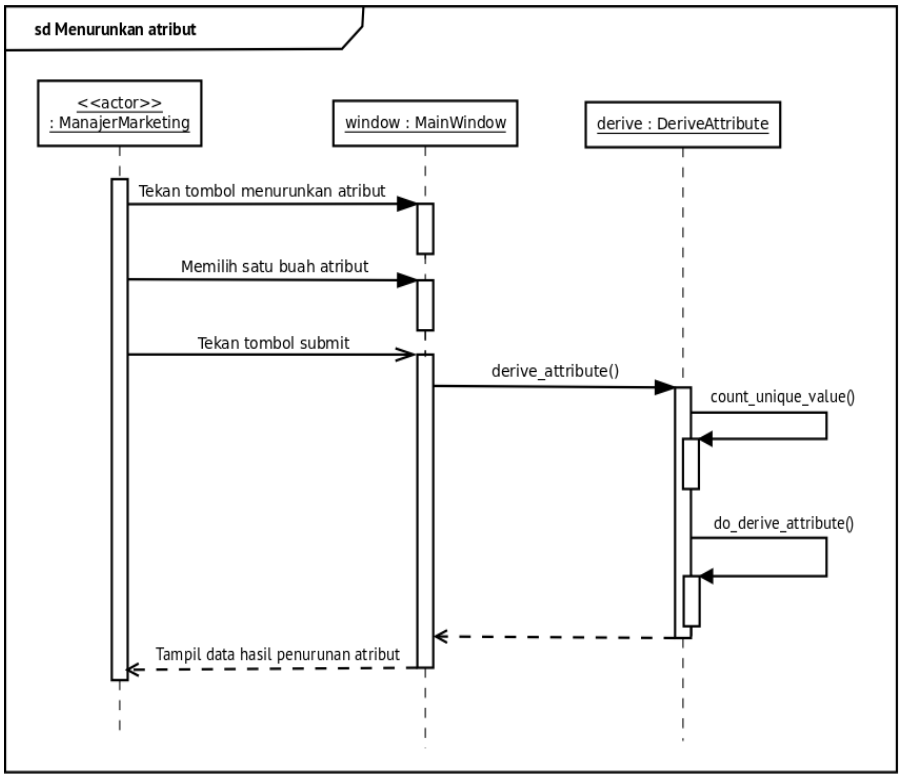
Sequence Diagram ini menggambarkan urutan interaksi antar objek pada proses *binning* atribut yang akan digunakan untuk pemodelan. *Sequence Diagram* binning atribut dapat dilihat pada Gambar III.22.



Gambar III.22 *Sequence diagram* Binning Atribut

5. *Sequence Diagram* Menurunkan Atribut

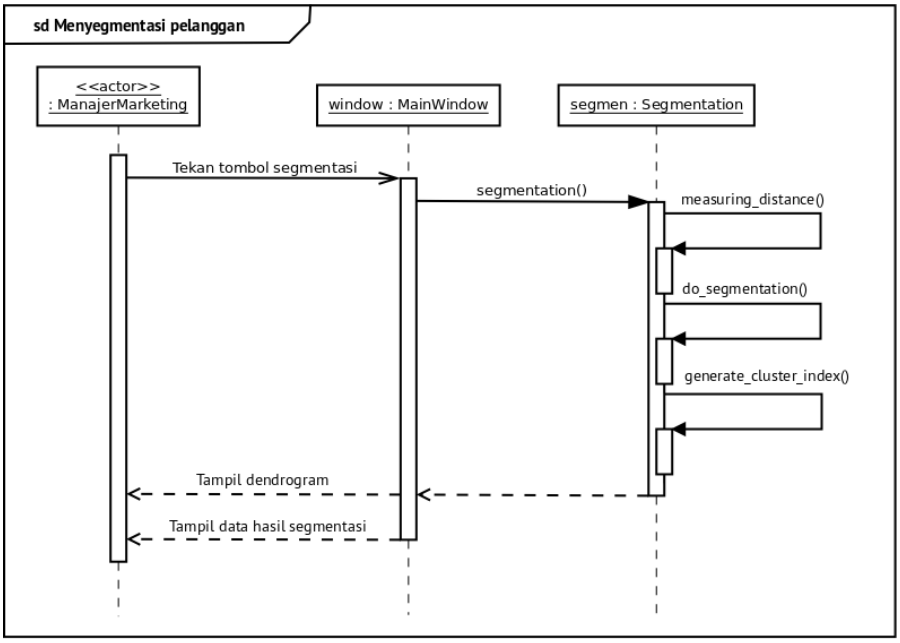
Sequence Diagram ini menggambarkan urutan interaksi antar objek pada proses penurunan atribut yang akan digunakan untuk pemodelan. *Sequence Diagram* menurunkan atribut dapat dilihat pada Gambar III.23.



Gambar III.23 *Sequence diagram* Menurunkan Atribut

6. *Sequence Diagram* Menyegmentasi Pelanggan

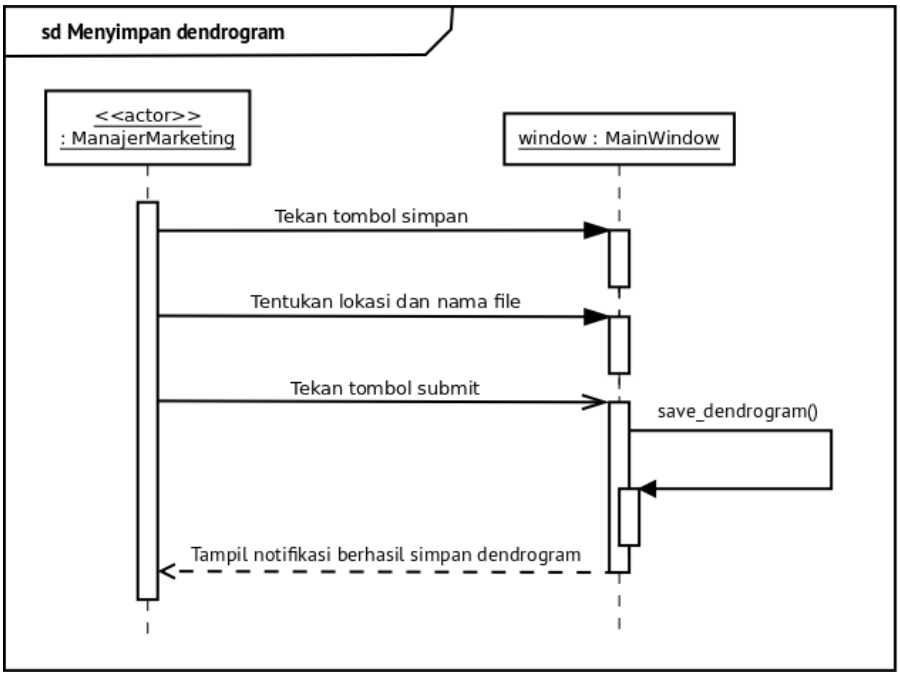
Pada diagram ini digambarkan interaksi objek-objek pada proses segmentasi. *Sequence Diagram* menyegmentasi pelanggan disajikan pada Gambar III.24.



Gambar III.24 *Sequence diagram* menyegmentasi pelanggan

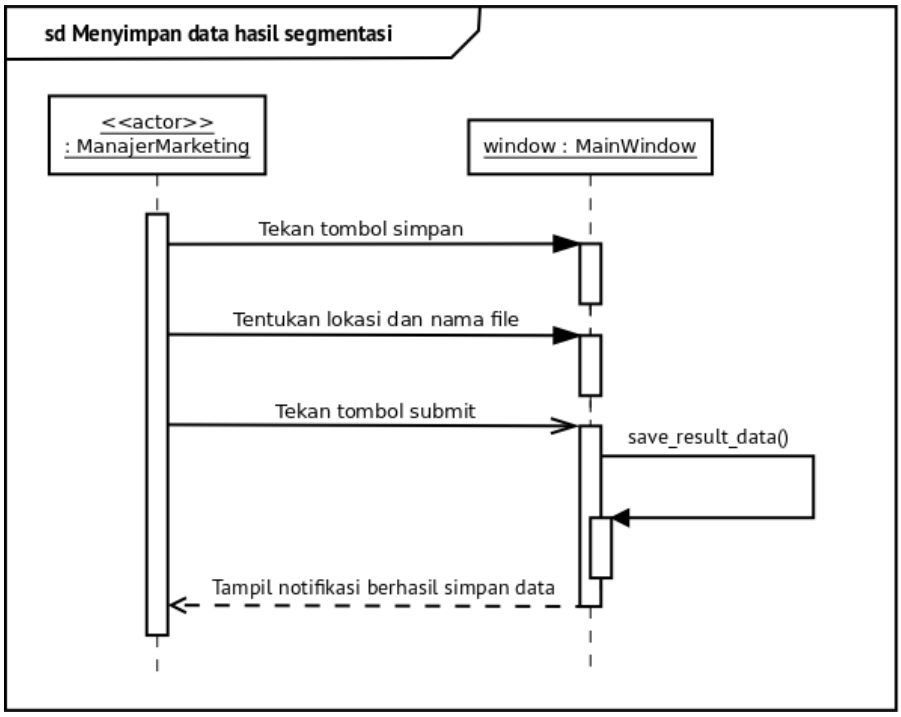
7. *Sequence Diagram* Menyimpan Hasil Segmentasi

Sequence Diagram menyimpan *dendrogram* disajikan pada Gambar III.25.



Gambar III.25 *Sequence diagram* menyimpan *dendrogram*

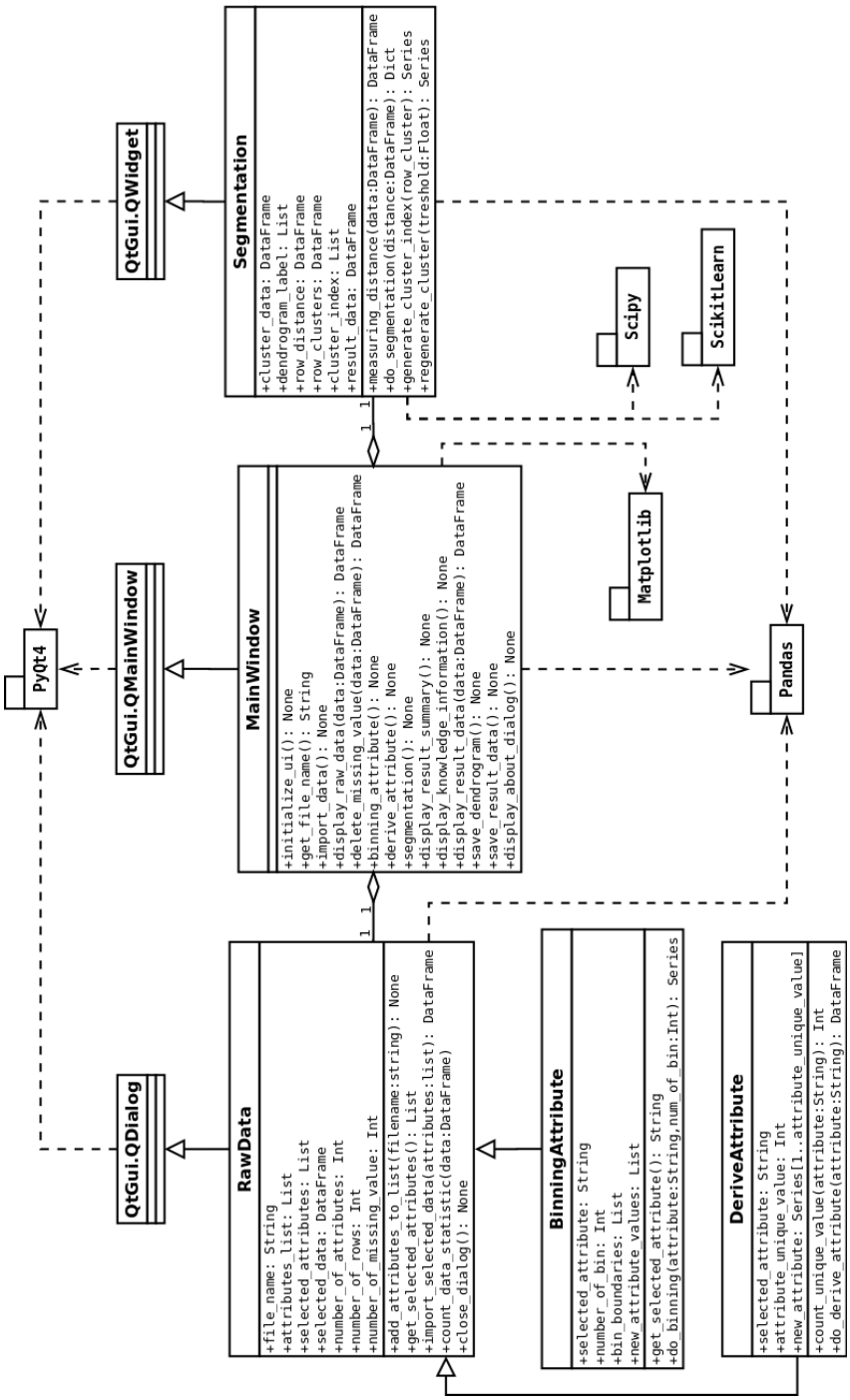
Sedangkan *Sequence Diagram* menyimpan data hasil segmentasi dapat dilihat pada Gambar III.26.



Gambar III.26 *Sequence diagram* menyimpan data hasil segmentasi

III.1.8.4. Class Diagram

Class diagram dibuat untuk menggambarkan tipe dan perilaku objek dalam sistem serta hubungan antar setiap objek. Gambar III.27 merupakan *class diagram* dari sistem segmentasi pelanggan.



Gambar III.27 *Class diagram* sistem segmentasi pelanggan

III.2. Perancangan Sistem

Tahap perancangan sistem merupakan tahapan pembuatan desain sistem yang akan dibangun. Kegiatan pada proses perancangan sistem meliputi perancangan

data, perancangan struktur menu, perancangan antarmuka, dan perancangan jaringan semantik.

III.2.1. Perancangan Data

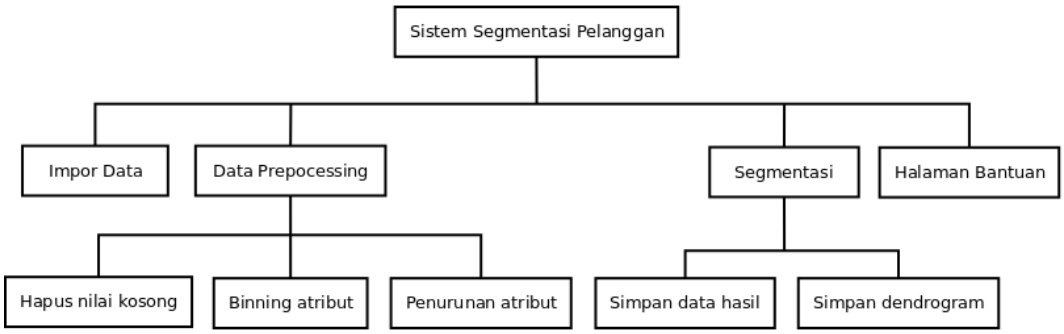
Tahap perancangan data digunakan untuk menjelaskan karakteristik data yang akan diolah oleh sistem. Struktur tabel digunakan untuk menjelaskan rincian data pada tabel yang meliputi nama atribut, tipe data, panjang data, dan keterangan lainnya. Tabel III.32 merupakan struktur tabel pelanggan pada basis data sistem segmentasi pelanggan.

Tabel III.32 Struktur tabel pelanggan pada sistem segmentasi pelanggan

Nama atribut	Tipe data	Panjang	Keterangan
id_pelanggan	varchar	8	Unique
nama	varchar	50	-
alamat	varchar	100	-
kota	varchar	25	-
provinsi	varchar	25	-
no_telepon	varchar	30	-
item	varchar	200	-
custom_name	char	1	-
jumlah_item	integer	3	-
jumlah_transaksi	integer	2	-
total_transaksi	integer	8	-
metode_pembayaran	varchar	10	-

III.2.2. Perancangan Struktur Menu

Struktur menu perlu dibentuk untuk memudahkan alur penggunaan sistem oleh pengguna. Perancangan menu yang digunakan untuk sistem segmentasi pelanggan adalah struktur menu hierarki. Terdapat empat menu utama pada sistem yaitu impor data pelanggan, *data preprocessing,* segmentasi pelanggan, dan halaman bantuan. Gambar III.28 merupakan perancangan struktur menu sistem segmentasi pelanggan.



Gambar III.28 Struktur menu sistem segmentasi pelanggan

III.2.3. Perancangan Antarmuka Perangkat Lunak

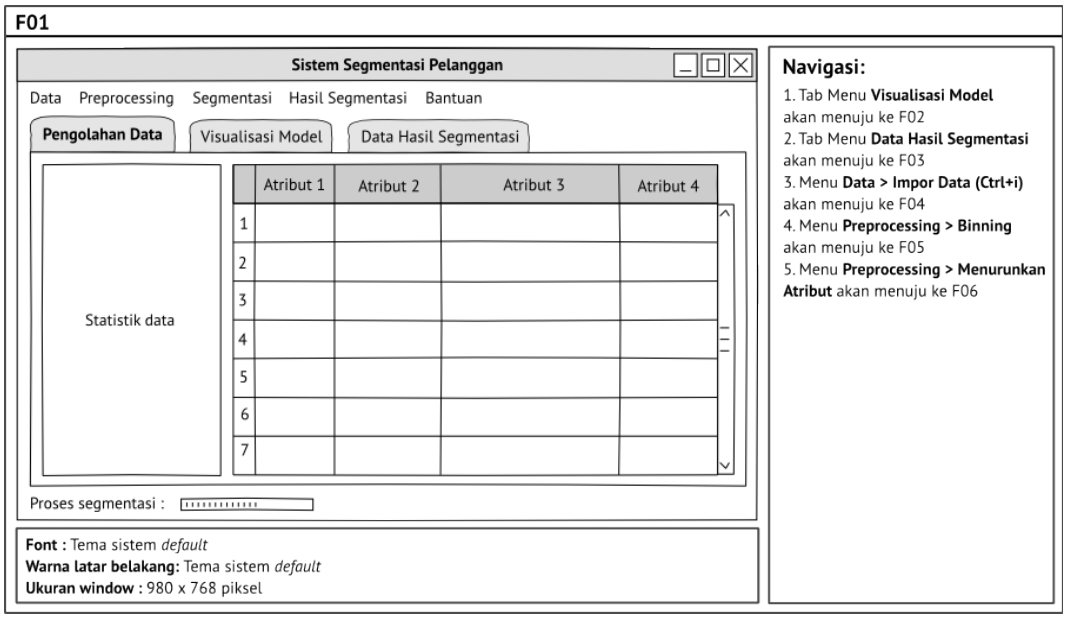
Perancangan antarmuka perangkat lunak menggambarkan desain tampilan dari sistem segmentasi pelanggan. Sesuai dengan spesifikasi kebutuhan non-fungsional pada Tabel III.18, antarmuka dari sistem harus sederhana. Hal tersebut akan dapat memudahkan pengguna dalam mengoperasikan sistem segmentasi pelanggan. Perancangan antarmuka meliputi perancangan *form* dan perancangan pesan.

III.2.3.1. Perancangan Form

Perancangan *form* dibutuhkan sebagai gambaran *window* dan tata letak berbagai *widget* pada sistem segmentasi pelanggan. Pada tahap ini akan dibuat rancangan GUI (Graphical User Interface) yang akan diimplementasikan dengan *Qt Toolkit* [16]. Beberapa fitur yang akan digunakan adalah *titlebar*, *menubar*, *toolbar*, *statusbar*, dan *shortcut*. Perancangan *form* pada sistem segmentasi pelanggan adalah sebagai berikut.

a. Form pengolahan data

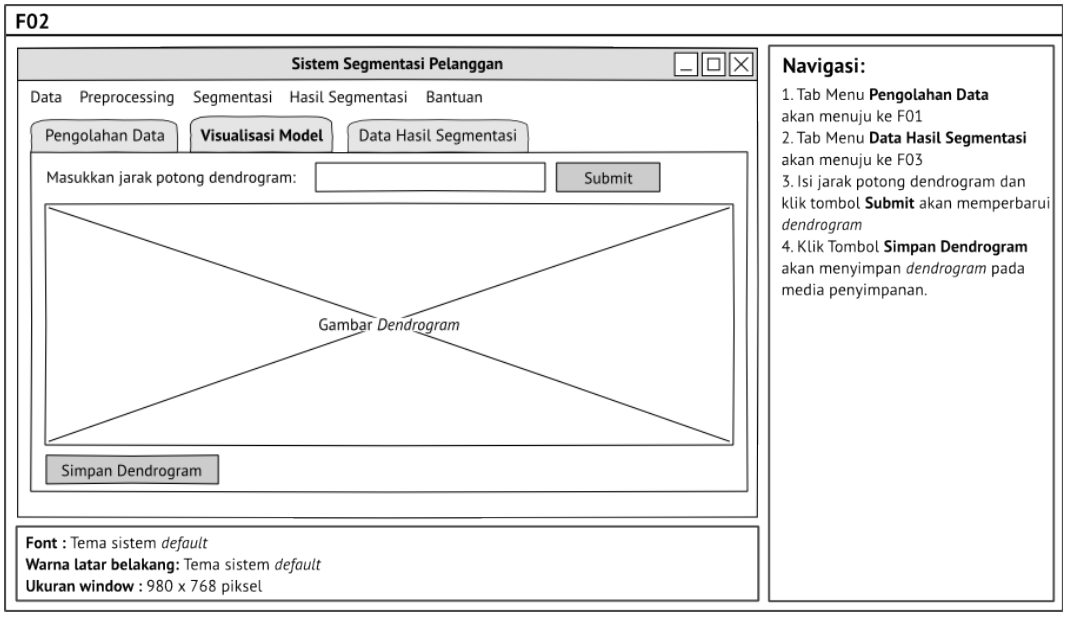
Gambar III.29 merupakan perancangan *form* pengolahan data pada sistem segmentasi pelanggan.



Gambar III.29 Perancangan *form* pengolahan data

b. Form visualisasi model

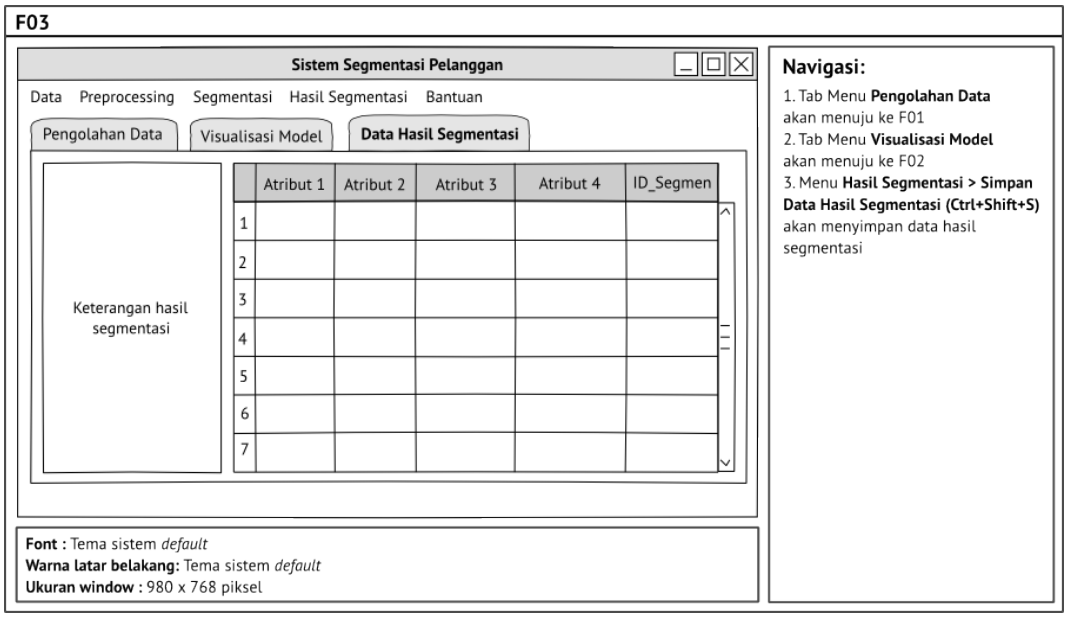
Gambar III.30 merupakan perancangan *form* untuk visualisasi model hasil segmentasi pada sistem segmentasi pelanggan yang akan dibangun.



Gambar III.30 Perancangan *form* visualisasi model

c. Form menampilkan data hasil segmentasi pelanggan

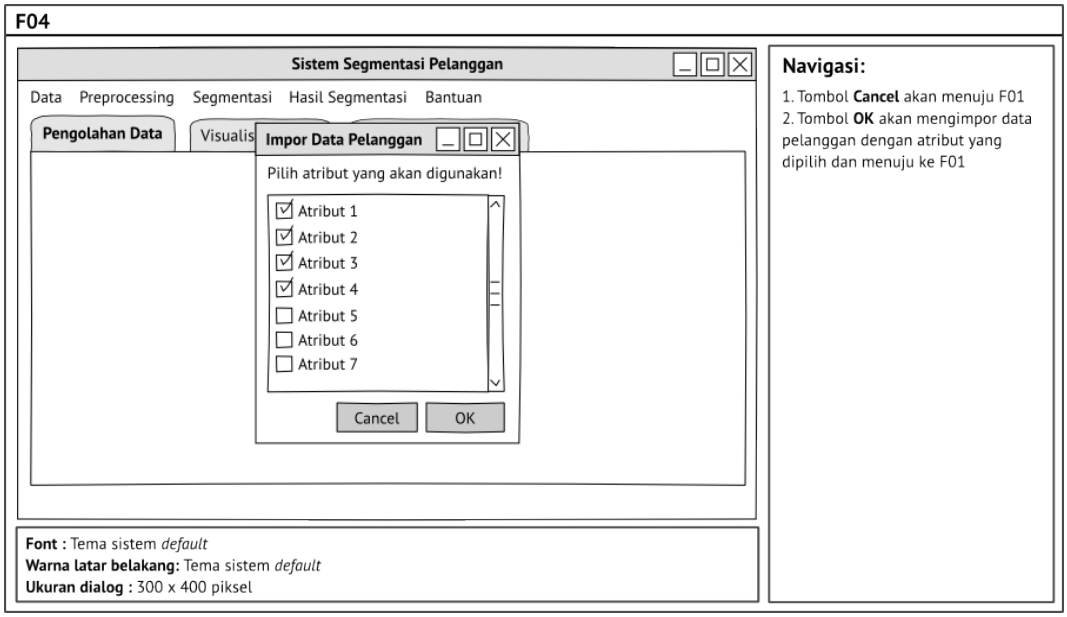
Gambar III.31 merupakan perancangan *form* untuk menampilkan data hasil segmentasi pelanggan pada sistem segmentasi pelanggan yang akan dibangun.



Gambar III.31 Perancangan *form* menampilkan data hasil segmentasi

d. Form mengimpor data pelanggan

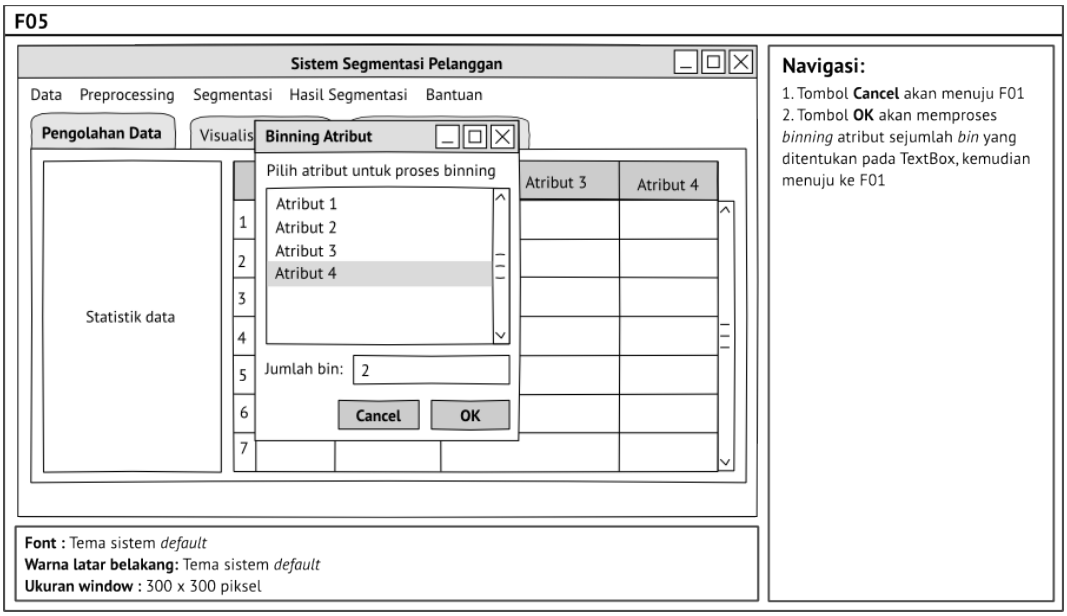
Gambar III.32 merupakan perancangan *form* untuk mengimpor data pelanggan pada sistem segmentasi pelanggan yang akan dibangun.



Gambar III.32 Perancangan *form* mengimpor data pelanggan

e. Form *binning* atribut

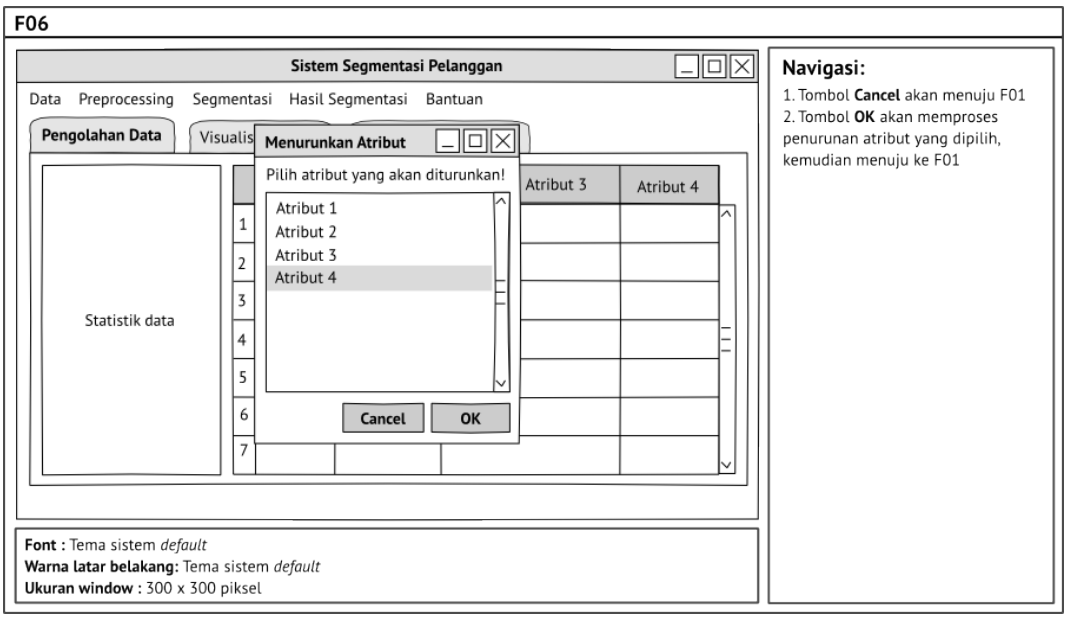
Gambar III.33 merupakan perancangan *form* untuk *binning* atribut pada sistem segmentasi pelanggan yang akan dibangun.



Gambar III.33 Perancangan *form* binning atribut

f. Form menurunkan atribut

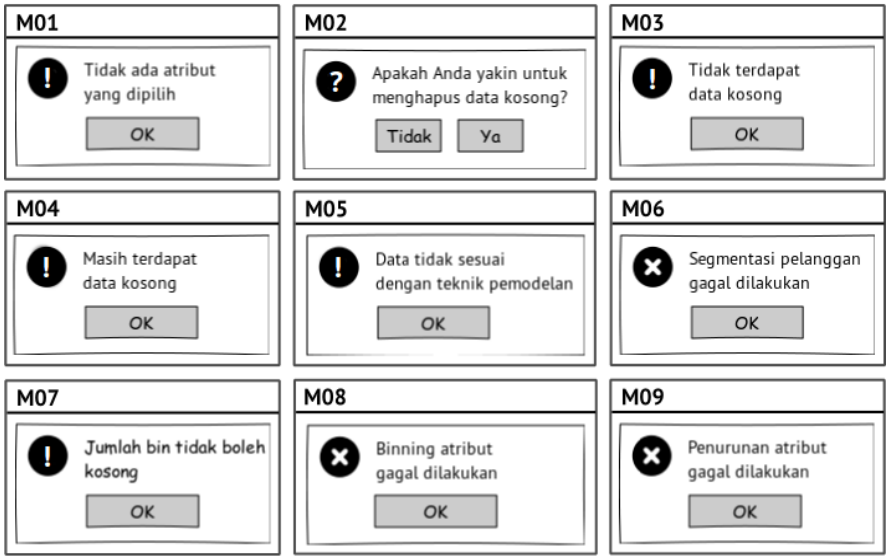
Gambar III.34 merupakan perancangan *form* untuk menurunkan atribut pada sistem segmentasi pelanggan yang akan dibangun.



Gambar III.34 Perancangan *form* menurunkan atribut

III.2.3.2. Perancangan Pesan

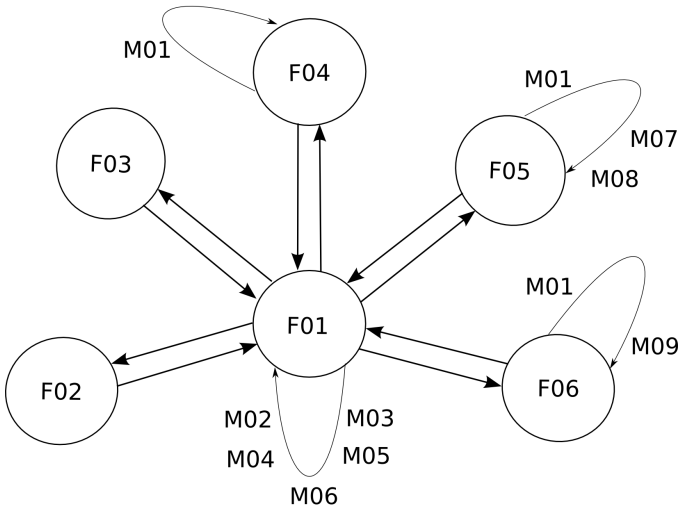
Pesan pada sistem segmentasi pelanggan dirancang dalam bentuk dialog. Pesan tersebut meliputi pesan peringatan, pesan konfirmasi, dan pesan kesalahan. Gambar III.35 merupakan perancangan pesan pada sistem segmentasi pelanggan.



Gambar III.35 Perancangan pesan sistem segmentasi pelanggan

III.2.4. Jaringan Semantik

Jaringan semantik menggambarkan hubungan antara komponen-komponen antarmuka yang terdiri *form* dan pesan. Gambar III.36 merupakan jaringan semantik dari sistem segmentasi pelanggan.



Gambar III.36 Jaringan semantik sistem segmentasi pelanggan