



TUGAS AKHIR - KS151601

**SEGMENTASI PELANGGAN MENGGUNAKAN
CLUSTERING K-MEANS DAN MODEL RFM (STUDI
KASUS: PT. BINA ADIDAYA SURABAYA)**

***CUSTOMER SEGMENTATION USING K-MEANS
CLUSTERING AND RFM MODEL (CASE STUDY:
PT. BINA ADIDAYA SURABAYA***

Anissa Veronika Angelie
NRP 5213 100 071

Dosen Pembimbing:
Edwin Riksakomara, S.Kom., M.T.

DEPARTEMEN SISTEM INFORMASI
Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember
Surabaya 2017

TUGAS AKHIR - KS151601

**SEGMENTASI PELANGGAN per MENGGUNAKAN
CLUSTERING K-MEANS DAN MODEL RFM (STUDI
KASUS: PT. BINA ADIDAYA SURABAYA)**

Anissa Veronika Angelle

NRP 5213 100 071

Dosen Pembimbing:
Edwin Riksakomara, S.Kom., M.T.

DEPARTEMEN SISTEM INFORMASI
Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember
Surabaya 2017

FINAL PROJECT - KS151601

**CUSTOMER SEGMENTATION USING K-MEANS
CLUSTERING AND RFM MODEL (CASE STUDY:
PT. BINA ADIDAYA SURABAYA**

Anissa Veronika Angelie

NRP 5213 100 071

Supervisor:
Edwin Riksakomara, S.Kom., M.T.

DEPARTMENT OF INFORMATION SYSTEMS
Faculty of Information Technology
Sepuluh Nopember Institute of Technology
Surabaya 2017

LEMBAR PENGESAHAN

**SEGMENTASI PELANGGAN MENGGUNAKAN
CLUSTERING K-MEANS DAN MODEL RFM (STUDI
KASUS: PT. BINA ADIDAYA SURABAYA)**

TUGAS AKHIR

Disusun untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer
pada
Departemen Sistem Informasi
Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember

Oleh:

ANISSA VERONIKA ANGELIE
5213 100 071

Surabaya, Juli 2017

KEPALA
DEPARTEMEN SISTEM INFORMASI

Dr.Ir. Aris Tjahyantq, M.Kom.
NIP/19650310 199102 1 001

LEMBAR PERSETUJUAN

SEGMENTASI PELANGGAN MENGGUNAKAN *CLUSTERING* K-MEANS DAN MODEL RFM (STUDI KASUS: PT. BINA ADIDAYA SURABAYA)

TUGAS AKHIR

Disusun untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer
pada
Departemen Sistem Informasi
Fakultas Teknologi Informasi
Institut Teknologi Sepuluh Nopember

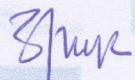
Oleh :

ANISSA VERONIKA ANGELIE

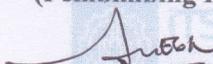
5213 100 071

Disetujui Tim Penguji : Tanggal Ujian : 4 Juli 2017
Periode Wisuda : September 2017

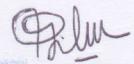
Edwin Riksakomara, S.Kom., M.T.


(Pembimbing I)

Wiwik Anggraeni, S.Si, M.Kom.


(Penguji I)

Renny Pradina K, S.T., M.T.


(Penguji II)

**SEGMENTASI PELANGGAN MENGGUNAKAN
CLUSTERING K-MEANS DAN MODEL RFM (STUDI
KASUS: PT. BINA ADIDAYA SURABAYA)**

Nama Mahasiswa : Anissa Veronika Angelie
NRP : 5213 100 071
Departemen : Sistem Informasi
Pembimbing 1 : Edwin Riksakomara, S.Kom, M.T

ABSTRAK

Pada masa Masyarakat Ekonomi ASEAN (MEA), Indonesia dipilih oleh beberapa perusahaan dari negara lain untuk menjual produknya, termasuk diantaranya perusahaan cat luar negeri. Oleh karena itu, persaingan pasar yang semakin sengit membuat PT. Bina Adidaya Surabaya tidak mungkin hanya berfokus dengan produk yang dijual, namun penting juga untuk memperhatikan proses mengelolah hubungan pelanggan yaitu dengan retailer. Segmentasi merupakan proses awal yang mengetahui pelanggan mana yang dapat dipertahankan. Pada segmentasi, pelanggan yang memiliki kesamaan tertentu akan dikelompokkan menjadi satu untuk mempermudah proses pengelolaan.

Metode clustering dengan algoritma K-Means diterapkan untuk melakukan segmentasi. Metode Elbow digunakan untuk membantu menemukan jumlah k dalam penerapan K-Means. Implementasi clustering melibatkan variabel yang didapatkan dari model RFM. Model tersebut dapat menggambarkan perilaku retailer dalam melakukan transaksi dengan PT. Bina Adidaya. Data yang digunakan dalam segmentasi pelanggan merupakan riwayat pembelian retailer pada periode Januari 2016 hingga Desember 2016.

Penerapan analisis RFM dan metode K-Means menghasilkan dua macam segmen yaitu menjadi 2 segmen dan 3 segmen, namun 3 segmen pelanggan merupakan pembagian yang terbaik. Hasil 3 segmen telah diuji performanya menggunakan tiga metode yaitu performa berdasarkan Sum of Square Error sebesar 3.08, Dunn Index sebesar 0,25 dan Connectivity sebesar 7.89. Persentase jumlah anggota pada segmen 1 sebesar 9.64%, segmen 2 sebesar 65.1% dan segmen 3 sebesar 25.3%. Analisa berdasarkan nilai RFM setiap segmen menunjukkan bahwa segmen 1 merupakan segmen yang memiliki nilai terbaik. Tugas akhir ini menghasilkan visualisasi segmentasi pelanggan PT. Bina Adidaya Surabaya dengan grafik dan gambar yang interaktif pada web-based.

Kata Kunci : segmentasi pelanggan, clustering, model RFM, K-Means, Metode Elbow

**CUSTOMER SEGMENTATION USING K-MEANS
CLUSTERING AND RFM MODEL (CASE STUDY: PT.
BINA ADIDAYA SURABAYA**

Student Name : Anissa Veronika Angelie
NRP : 5213 100 071
Department : Sistem Informasi
Supervisor 1 : Edwin Riksakomara, S.Kom, M.T

ABSTRACT

At the time of the ASEAN Economic Community (MEA), Indonesia was selected by several companies from other countries to sell its products, including overseas paint companies. Therefore, the increasingly fierce market competition makes PT. Bina Adidaya Surabaya is unlikely to focus solely on products sold, but it is also important to pay attention to the process of managing customer relationships with retailers. Segmentation is an early process that knows which customers can be sustained. In segmentation, customers who have certain similarities will be grouped into one for.

The grouping method with K-Means algorithm is applied to segment. The Elbow method is used to help find the number of k in the application of K-Means. Implementation Clustering involves variables obtained from the RFM model. The model can describe the behavior of retailers in conducting transactions with PT. Bina Adidaya. The data used in customer segmentation is the history of retailer purchases in the period January 2016 through December 2016.

The application of RFM and K - Means method resulted in two segments: 2 segments and 3 segments, but 3 segment customers is the best division. The 3 segment results have been tested for

performance using three methods: performance based on Sum of Square Error is 3.08, Dunn Index is 0.25 and Connectivity is 7.89. Percentage number of members at segment number 1 is 9.6%, segment 2 is 65.1% and segment 3 is 25.3%. The analysis based on the RFM value of each segment shows that segment 1 is a segment that has the best value. This final project resulted in visualization of customer segmentation of PT. Bina Adidaya Surabaya with interactive graphics and images on the web-based.

Keywords : customer segmentation, clustering, RFM Model, K-Means, Elbow Method

KATA PENGANTAR

Puji dan syukur penulis tuturkan kepada Allah SWT karena atas kasih karunia dan pertolongan-Nya, penulis dapat menyelesaikan laporan tugas akhir dengan judul: “**Segmentasi Pelanggan Menggunakan *Clustering* K-Means dan Model RFM (Studi Kasus: PT. Bina Adidaya Surabaya)**” yang merupakan salah satu karya penulis sekaligus syarat kelulusan pada Departemen Sistem Informasi, Fakultas Teknologi Informasi, Institut Teknologi September Sepuluh Nopember Surabaya.

Tugas akhir ini tidak akan pernah terwujud tanpa bantuan dan dukungan dari berbagai pihak yang sudah meluangkan waktu, tenaga, dan pikirannya bagi penulis dalam menyelesaikan tugas akhir ini. Pada kesempatan ini penulis ingin mengucapkan terima kasih sebesar-besarnya kepada pihak di bawah ini, yaitu:

1. Allah SWT yang telah memberikan kesehatan, kemudahan dan kesempatan untuk bisa menyelesaikan tugas akhir ini.
2. Kedua orang tua, adik dan keluarga yang senantiasa selalu mendoakan dan memberikan kasih sayang serta semangat tiada henti untuk menyelesaikan Tugas Akhir.
3. Kepada pihak PT. Bina Adidaya cabang Surabaya yang telah memberikan kesempatan untuk meleakukan penelitian dan bantuan selama proses pengerjaan tugas akhir.
4. Bapak Edwin Riksakomara, S.Kom., M.T., selaku dosen pembimbing yang telah banyak meluangkan waktu untuk membimbing, mengarahkan, dan mendukung penulis dalam penyelesaian Tugas Akhir.
5. Bapak Bakti Cahyo Hidayanto, S.Si., M.Kom. selaku dosen wali yang senantiasa memberikan pengarahan dan motivasi selama penulis menempuh masa perkuliahan dan pengerjaan Tugas Akhir.
6. Mahendrawati E.R. S.T., M.Sc., Ph.D., Wiwik Anggraeni, S.Si, M.Kom., Renny Pradina K, S.T., M.T., Bapak dan Ibu dosen pengajar seluruhnya beserta karyawan Departemen Sistem Informasi, FTIF ITS Surabaya yang telah

memberikan ilmu, pengalaman berharga dan bantuan kepada penulis selama masa perkuliahan

7. Canal, Gajah (Ofi, Nena, Izza, Mira, Risa, Hemas, Amal, Pecol) dan teman-teman seperjuangan BELTRANIS semuanya, yang selalu memberikan semangat positif untuk menyelesaikan Tugas Akhir dengan tepat waktu.

Penulis menyadari bahwa Tugas Akhir ini masih belum sempurna dan memiliki banyak kekurangan di dalamnya. Oleh karena itu, penulis membuka pintu selebar-lebarnya bagi pihak yang ingin memberikan kritik dan saran. Semoga buku Tugas Akhir ini bermanfaat bagi pihak yang membutuhkan.

Surabaya, Juli 2017

Penulis

DAFTAR ISI

ABSTRAK.....	i
ABSTRACT.....	iii
KATA PENGANTAR	v
DAFTAR ISI.....	vii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xiii
DAFTAR SCRIPT	xiv
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Perumusan Masalah	4
1.3. Batasan Masalah	4
1.4. Tujuan Tugas Akhir	5
1.5. Manfaat Tugas Akhir	5
1.6. Relevansi Tugas Akhir.....	5
1.7. Sistematika Penulisan Tugas Akhir	5
BAB II TINJAUAN PUSTAKA.....	9
2.1. Studi Sebelumnya	9
2.2. Dasar Teori.....	13
2.2.1. PT. Bina Adidaya Cabang Surabaya.....	13
2.2.2. Pengelolaan Hubungan Pelanggan	15
2.2.3. Segmentasi Pelanggan.....	16
2.2.4. Penggalian Data.....	17
2.2.5. Clustering	18
2.2.6. Algoritma K-Means.....	20
2.2.7. Metode Elbow	21
2.2.8. Normalisasi Min-Max	22
2.2.9. Sum of Square Error (SSE)	23
2.2.10. Dunn Index.....	23
2.2.11. Connectivity	24
2.2.12. Model RFM.....	24
2.2.13 Pembobotan Menggunakan Analytical Hierarchy Process (AHP).....	26
2.2.14. Customer Lifetime Value (CLV)	28
BAB III METODOLOGI PENELITIAN.....	29
3.1. Metodologi.....	29

3.1.1	Identifikasi Masalah	30
3.1.2	Studi Literatur	30
3.1.3	Pengumpulan Data.....	30
3.1.4	Pra Proses Data.....	31
3.1.5	Proses Clustering	31
3.1.6	Verifikasi Hasil Clustering	32
3.1.7	Uji Performa Hasil Clustering	32
3.1.8	Pemberian Bobot RFM.....	32
3.1.9	Perhitungan Customer Lifetime Value (CLV)	32
3.1.10	Pembuatan Visualisasi	33
3.1.11	Analisis Hasil Segmentasi Pelanggan.....	33
3.1.12	Penarikan kesimpulan dan Saran.....	33
3.1.13	Penyusunan Laporan Tugas Akhir.....	33
BAB IV	PERANCANGAN.....	35
4.1.	Pengumpulan Data	35
4.2.	Pra Proses Data.....	38
4.3.	Pemberian Bobot RFM.....	39
4.4.	Konstruksi Visualisasi.....	39
4.4.1.	Use Case Diagram	39
4.4.2.	Deskripsi Use Case.....	41
4.4.3.	Desain Antarmuka Visualisasi.....	42
BAB V	IMPLEMENTASI	45
5.1.	Pra Proses Data.....	45
5.1.1.	Pemilihan Data	45
5.1.2.	Pembersihan Data.....	45
5.1.3.	Integrasi Data.....	46
5.1.4.	Transformasi Data	48
5.2.	Proses Clustering.....	50
5.2.1.	Penentuan Nilai k dengan Metode Elbow	51
5.2.2.	Clustering dengan K-Means	51
5.3.	Verifikasi Hasil Clustering.....	52
5.4.	Uji Performa Clustering	53
5.5.	Pemberian Bobot RFM.....	54
5.6.	Perhitungan CLV.....	56
5.7.	Pembuatan Visualisasi dengan RShiny	58
5.7.1.	Halaman 2 Segmen dan 3 Segmen	58
5.7.2.	Halaman Detail 2 dan 3 Segmen	60
5.7.3.	Halaman Tabel data	65

BAB VI HASIL PEMBAHASAN	67
6.1. Proses Clustering	67
6.1.1. Penentuan Nilai k dengan Metode Elbow	67
6.1.2. Clustering dengan K-Means	68
6.2. Verifikasi Hasil Clustering	69
6.2.1. Verifikasi 2 Segmen	69
6.2.2. Verifikasi 3 Segmen	70
6.3. Uji Performa Clustering	71
6.4. Pemberian Bobot RFM	73
6.5. Perhitungan CLV	74
6.5.1. Perhitungan CLV 2 Segmen	74
6.5.2. Perhitungan CLV 3 Segmen	75
6.6. Analisis Cluster yang Terbentuk	75
6.6.1. Analisis Karakteristik 2 Segmen	75
6.6.2. Analisis Karakteristik 3 Segmen	76
6.6.3. Analisis Persamaan dan Perbedaan 2 Segmen dan 3 Segmen	79
6.7. Analisis berdasarkan Visualisasi	79
6.7.1. Analisis Visualisasi 2 Segmen	79
6.7.2. Analisis Visualisasi 3 Segmen	84
6.8. Kesimpulan Analisis	88
BAB VII KESIMPULAN DAN SARAN	91
7.1. Kesimpulan	91
7.2. Saran	92
DAFTAR PUSTAKA	93
BIODATA PENULIS	99
A. Lampiran A	A-1
B. Lampiran B	B-1
C. Lampiran C	C-1
D. Lampiran D	D-1
E. Lampiran E	E-1
F. Lampiran F	F-1

Halaman ini sengaja dikosongkan

DAFTAR TABEL

Tabel 2.1 Studi sebelumnya judul paper 1	9
Tabel 2.2 Studi sebelumnya judul paper 2	10
Tabel 2.3 Studi sebelumnya judul paper 3	11
Tabel 2.4 Studi sebelumnya judul paper 4	12
Tabel 2.5 Studi sebelumnya judul paper 5	12
Tabel 2.6 Preference Scale	26
Tabel 2.7 Random Indeks.....	28
Tabel 4.1 Potongan data retailer pelanggan PT. Bina Adidaya area Surabaya	36
Tabel 4.2 Potongan laporan pengeluaran barang berdasarkan surat jalan	36
Tabel 4.3 Potongan laporan berdasarkan surat jalan dan invoice	37
Tabel 4.4 Use case description “Melihat pembagian jumlah anggota pada grafik pie chart dan bar chart”.....	41
Tabel 5.1 Pemilihan atribut	46
Tabel 5.2 Baris data yang memiliki nilai biaya nol.....	46
Tabel 5.3 Nilai maksimal dan minimal setiap variabel	50
Tabel 5.4 Pairwise comparisons matrix	55
Tabel 5.5 Normal comparison matrix dan bobot.....	55
Tabel 5.6 Perhitungan overall score	55
Tabel 5.7 Nilai rata-rata RFM atau centroid 2 segmen	57
Tabel 5.8 Nilai rata-rata RFM atau centroid 3 segmen	57
Tabel 5.9 Peringkat CLV 2 segmen	57
Tabel 5.10 Peringkat CLV 3 segmen	57
Tabel 6.1 Potongan hasil perhitungan Euclidean Distance untuk 2 segmen.....	70
Tabel 6.2 Potongan hasil perhitungan Euclidean Distance untuk 3 segmen.....	70
Tabel 6.3 Rekapitulasi hasil uji performa clustering dari semua metode	73
Tabel 6.4 Bobot setiap kriteria	74
Tabel 6.5 CLV untuk jumlah segmen sebanyak 2.....	74
Tabel 6.6 CLV untuk jumlah segmen sebanyak 3.....	75
Tabel 6.7 Karakteristik cluster 1 ketika 2 segmen	76

Tabel 6.8 Karakteristik cluster 2 ketika 2 segmen.....	76
Tabel 6.9 Karakteristik cluster 1 ketika 3 segmen.....	77
Tabel 6.10 Karakteristik cluster 2 ketika 3 segmen.....	78
Tabel 6.11 Karakteristik cluster 3 ketika 3 segmen.....	78
Tabel A.1 Data tugas akhir hasil pra proses yang digabungkan dengan hasil clustering	A-1
Tabel B.1 Atribut data mentah	B-1
Tabel C.1 hasil perhitungan Euclidean Distance untuk 2 segmen	C-1
Tabel C.2 hasil perhitungan Euclidean Distance untuk 3 segmen	C-6
Tabel E.1 Use case description “Melihat persebaran retailer pada grafik 3 dimensi”.....	E-1
Tabel E.2 Use case description “Melihat persebaran retailer pada peta lokasi”.....	E-2
Tabel E.3 Use case description “Melihat peringkat retailer”.....	E-3
Tabel E.4 Use case description “Melihat persebaran variabel setiap segmen pada box plot”	E-4
Tabel E.5 Use case description “Melihat range setiap segmen”	E-5
Tabel E.6 Use case description “Melihat rata-rata nilai variabel setiap segmen”.....	E-6
Tabel E.7 Use case description “Melihar data tabel”	E-7

DAFTAR GAMBAR

Gambar 3.1 Metodologi tugas akhir.....	29
Gambar 4.1 Kriteria pemilihan.....	39
Gambar 4.2 Use case diagram.....	40
Gambar 4.3 Desain antarmuka halaman 2 dan 3 segmen.....	42
Gambar 4.4 Desain antarmuka halaman detail 2 dan 3 segmen	43
Gambar 4.5 Desain antarmuka halaman tabel data	43
Gambar 5.1 Halaman 2 segmen	59
Gambar 5.2 Halaman 3 segmen	60
Gambar 5.3 Halaman detail 2 segmen	65
Gambar 5.4 Halaman detail 3 segmen	65
Gambar 5.5 Halaman tabel data	66
Gambar 6.1 Grafik hasil penerapan metode Elbow	67
Gambar 6.2 Hasil clustering K-Means untuk 2 cluster	68
Gambar 6.3 Hasil clustering K-Means untuk 3 cluster	69
Gambar 6.4 Grafik hasil uji performa dengan metode SSE...	71
Gambar 6.5 Grafik hasil uji performa dengan metode Connectivity	72
Gambar 6.6 Grafik hasil uji performa dengan metode Dunn Index.....	72
Gambar 6.7 Grafik bar chart dan pie chart untuk 2 segmen...	80
Gambar 6.8 Grafik 3 dimensi untuk 2 segmen.....	80
Gambar 6.9 Peta lokasi untuk 2 segmen	81
Gambar 6.10 Peringkat retailer segmen 1 untuk 2 segmen....	81
Gambar 6.11 Box plot recency untuk 2 segmen	82
Gambar 6.12 Box plot frequency untuk 2 segmen.....	83
Gambar 6.13 Box plot monetary untuk 2 segmen.....	83
Gambar 6.14 Grafik bar chart dan pie chart untuk 3 segmen.	84
Gambar 6.15 Grafik 3 dimensi untuk 3 segmen.....	85
Gambar 6.16 Peta lokasi untuk 3 segmen	85
Gambar 6.17 Peringkat retailer dengan inputan semua segmen	86
Gambar 6.19 Box plot recency untuk 3 segmen	87
Gambar 6.20 Box plot frequency untuk 3 segmen.....	87
Gambar 6.21 Box plot monetary untuk 3 segmen.....	88

DAFTAR SCRIPT

Script 5.1 Integrasi pertama.....	47
Script 5.2 Integrasi kedua.....	47
Script 5.3 pengubahan nilai ke dalam bentuk RFM.....	49
Script 5.4 Normalisasi Min-Max.....	50
Script 5.5 Input dan inisiasi data set.....	51
Script 5.6 Metode Elbow.....	51
Script 5.7 Clustering 3 segmen menggunakan K-Means.....	52
Script 5.8 Euclidean Distance 3 segmen.....	53
Script 5.9 Metode Sum of Square Error.....	54
Script 5.10 Metode Conectivity dan Dunn Index.....	54
Script 5.11 Visualisasi dengan Bar Chart 3 segmen.....	58
Script 5.12 Visualisasi dengan pie chart 3 segmen.....	59
Script 5.13 Visualisasi dengan grafik 3 dimensi pada 2 segmen dan 3 segmen.....	61
Script 5.14 Visualisasi dengan peta lokasi pada 3 segmen dan 2 segmen.....	61
Script 5.15 Visualisasi dengan box plot 3 segmen.....	62
Script 5.16 Visualisasi dengan valuebox untuk peringkat retailer.....	63
Script 5.17 Visualisasi dengan tabel untuk peringkat retailer.....	63
Script 5.18 Visualisasi untuk rentang RFM.....	64
Script 5.19 Visualisasi untuk rata-rata RFM.....	64
Script 5.20 Visualisasi tabel data.....	66

BAB I

PENDAHULUAN

Pada bagian pendahuluan ini, akan dijelaskan mengenai latar belakang, rumusan permasalahan, batasan tugas akhir, tujuan dan manfaat yang dihasilkan, relevansi dan sistematika penyusunan dari Tugas Akhir ini.

1.1. Latar Belakang Masalah

Pada masa Masyarakat Ekonomi ASEAN (MEA), Indonesia dipilih oleh beberapa perusahaan dari negara lain untuk menjual produknya karena populasi Indonesia yang besar, termasuk diantaranya perusahaan cat luar negeri [1]. Peningkatan jumlah perusahaan yang ada saat ini, membuat persaingan global yang terus meningkat termasuk persaingan pasar dalam industri cat menjadi sangat sengit [2]. Peningkatan jumlah perusahaan juga menjadikan pemahaman mengenai pelanggan dibutuhkan dalam membangun model keputusan bisnis dan analisis data transaksi pelanggan untuk membantu perencanaan komersial [3]. Agar tidak kalah bersaing, perusahaan dalam negeri mengembangkan pasar dan menyesuaikan perkembangan yang terjadi saat ini dengan menerapkan strategi bisnis yang didapatkan dari menganalisis kebutuhan pelanggannya. Salah satu perusahaan cat dalam negeri telah mengembangkan sistem baru yang diperlukan pelanggan dalam proses transaksi [4]. Oleh karena itu, perusahaan juga harus memperhatikan kebutuhan pelanggannya selain berfokus dengan produk yang dijual.

Pemahaman mengenai pelanggan terdapat dalam pengelolaan hubungan pelanggan yang menggambarkan strategi komprehensif dalam proses memperoleh, mempertahankan dan bermitra dengan pelanggan tertentu untuk menciptakan nilai bagi perusahaan dan pelanggan [5]. Tugas terpenting perusahaan merupakan mencari pelanggan, namun mempertahankan pelanggan lebih penting karena kehilangan pelanggan berarti kehilangan seluruh aliran bisnis dengan

pelanggan [6]. Dalam lingkungan yang dinamis perusahaan yang kompetitif harus mencoba untuk memperoleh wawasan mengenai kebutuhan, sikap, dan perilaku pelanggan yang akan dipertahankan. Namun pada kenyataannya, perusahaan pasti memiliki pelanggan yang perilakunya berbeda-beda dalam melakukan transaksi [7]. Idealnya setiap pelanggan harus diperlakukan sebagai individu dan perusahaan harus memperhatikan pelanggan satu persatu, akan tetapi pendekatan ini jelas tidak mungkin. Oleh karena itu, alternatif yang efisien adalah melakukan segmentasi pelanggan [8].

Segmentasi pelanggan merupakan sebuah model yang dibangun dalam mengelompokkan pelanggan sesuai dengan standar tertentu untuk digunakan menjadi variabel pengelompokan. Pelanggan akan menjadi satu kelompok yang sama apabila memiliki kesamaan tertentu, sementara kelompok atau segmen yang berbeda merupakan pelanggan memiliki karakteristik yang berbeda [9]. Setelah melakukan analisis perilaku pelanggan dengan segmentasi, maka dapat dilakukan perencanaan proses pemasaran di masa depan dalam mengirim pesan yang paling tepat untuk setiap segmen pelanggan [10].

PT Bina Adidaya di Surabaya merupakan salah satu perusahaan industri cat yang perlu mengelola hubungan dengan pelanggannya. Perusahaan ini tidak langsung menjual produknya kepada *end customer* melainkan menjualnya kepada *retailer*. Oleh karena itu, *retailer* yang menjadi pelanggan perusahaan ini sangat berperan penting dalam membantu menyalurkan produk perusahaan kepada *end customer*. Mempertahankan *retailer* yang ada saat ini menjadi sangat penting bagi PT. Bina Adidaya untuk dapat memperbaiki strategi pemasaran ke *retailer* dan memilih *retailer* yang sesuai untuk melakukan kolaborasi pemasaran. Perusahaan ini telah berdiri sejak 25 tahun yang lalu, sehingga memiliki pelanggan yang telah tersebar di Indonesia khususnya di Surabaya. Namun tidak semua *retailer* yang tercatat sebagai pelanggan merupakan *retailer* yang selalu melakukan transaksi di PT. Bina Adidaya. Perusahaan ini belum pernah menganalisis perilaku *retailer* dalam melakukan pembelian, sehingga belum

diterapkan strategi yang tepat dalam mempertahankan *retailer* tersebut. Kondisi terkini, perusahaan masih memberikan strategi layanan yang sama kepada semua *retailer*. Strategi yang dilakukan perusahaan misalnya, tidak membebaskan biaya pengiriman bagi *retailer* di pulau Jawa dan Bali, memberikan promo yang sama kepada seluruh *retailer* dan memberikan poin setiap pembelian produk. Hubungan antara *retailer* dengan perusahaan juga sangat bergantung pada komunikasi dari sales yang bertanggung jawab dan bersifat subjektif. Dampak buruknya jika perusahaan tidak memperbaiki pengolahan hubungan dengan *retailer*, maka perusahaan dapat kalah bersaing dengan kompetitor karena tidak dapat memberikan produk dan pelayanan yang lebih sesuai dengan kebutuhan pelanggannya.

Permasalahan yang dialami oleh PT. Bina Adidaya dalam mengelola hubungan pelanggan dapat diselesaikan dengan proses segmentasi yaitu penggalan data riwayat pembelian *retailer* pada periode tertentu. *Retailer* akan dikelompokkan menjadi beberapa segmen yang dibedakan berdasarkan perilaku *retailer* dalam melakukan pembelian. Perilaku *retailer* ini dapat digambarkan melalui model RFM yaitu *Recency*, *Frequency*, *Monetary*. Metode yang digunakan dalam segmentasi pelanggan yaitu dengan *clustering* algoritma K-Means. Variabel untuk melakukan *clustering* ada tiga, sesuai dengan hasil analisis model RFM. *Clustering* merupakan metode yang dapat digunakan untuk pengelompokan berdasarkan kemiripan dari variabel RFM, jika memiliki karakteristik yang sama akan dikelompokkan dalam satu *cluster* sedangkan jika karakteristiknya berbeda akan berada pada *cluster* yang lain [11].

Algoritma K-Means merupakan suatu algoritma *clustering* untuk mempartisi setiap data set hanya ke dalam satu *cluster*. Algoritma K-Means memiliki kemudahan untuk interpretasi, implementasi yang sederhana, kecepatan dalam konvergensi dan dapat beradaptasi, namun mempunyai masalah sensitifitas terhadap penentuan partisi awal jumlah *cluster* [12]. Metode Elbow dapat dilakukan untuk memperbaiki kelemahan dari

metode K-Means dengan cara memberikan visual berupa grafik yang menunjukkan titik banyaknya *cluster* dan perhitungan nilai k berdasarkan *Sum of Square Error (SSE)* [13].

Dengan adanya penelitian ini, diharapkan dapat memberikan kontribusi terhadap bagian pemasaran PT. Bina Adidaya karena analisa data dan visualisasi berbasis web-based akan membantu merencanakan strategi sesuai dengan perilaku setiap *retailer*.

1.2. Perumusan Masalah

Berdasarkan latar belakang yang telah dijabarkan di atas, berikut merupakan rumusan masalah yang menjadi acuan dalam tugas akhir ini:

1. Bagaimana menerapkan model RFM untuk melakukan segmentasi pelanggan PT. Bina Adidaya Surabaya?
2. Bagaimana visualisasi yang dapat membantu mempermudah pemahaman terhadap hasil segmentasi?

1.3. Batasan Masalah

Pada pengerjaan tugas akhir ini terdapat batasan mengenai area penelitian yang dilakukan yaitu:

1. Studi kasus tugas akhir yaitu PT. Bina Adidaya Cabang Surabaya dengan *retailer* yang berada pada area Surabaya.
2. Data yang digunakan yaitu data transaksi penjualan produk PT. Bina Adidaya di kota Surabaya dari rentang waktu 1 tahun yaitu pada bulan Januari 2016 hingga bulan Desember 2016
3. Kategori produk yang digunakan adalah produk pada jenis pemasaran retail.
4. Variabel yang digunakan dalam tugas akhir ini adalah waktu terakhir pembelian, frekuensi pembelian dan besarnya biaya pembelian.
5. Metode Elbow digunakan untuk membantu mencari masukan nilai k ketika proses *clustering* K-Means

1.4. Tujuan Tugas Akhir

Tujuan dari pengerjaan tugas akhir ini untuk membuat segmen pelanggan PT. Bina Adidaya dan mendeskripsikan karakteristik setiap segmen berdasarkan informasi pada hasil *clustering* dengan model RFM. Penelitian ini juga bertujuan memberikan tampilan hasil segmentasi pada visualisasi yang berbasis web-based untuk memudahkan membaca hasil penelitian.

1.5. Manfaat Tugas Akhir

Manfaat yang akan diperoleh bagi perusahaan dan peneliti berdasarkan tujuan yang diinginkan antara lain:

1. Bagi perusahaan, dapat membantu PT. Bina Adidaya cabang Surabaya dalam merencanakan strategi untuk mempertahankan *retailer* yang ada sesuai dengan perilaku *retailer* setiap segmennya.
2. Bagi peneliti, dapat menjadikan referensi mengenai model segmentasi pelanggan dan pembuatan visualisasi hasil segmentasi pada perusahaan yang memiliki jenis pelanggan *retailer*.

1.6. Relevansi Tugas Akhir

Relevansi tugas akhir ini terhadap penelitian area sistem informasi berada pada laboratorium Rekayasa Data dan Intelegensi Bisnis dengan topik penggalian data yaitu *clustering*. Penelitian ini merupakan penerapan dari beberapa mata kuliah seperti Sistem Cerdas, Penggalian Data dan Analitika Bisnis, dan Manajemen Rantai Pasok dan Hubungan Pelanggan.

1.7. Sistematika Penulisan Tugas Akhir

Sistematika penulisan laporan tugas akhir nantinya akan disusun sesuai dengan format pada Departemen Sistem Informasi. Penulisan ini dibagi menjadi tujuh bab dan setiap bab terdiri dari beberapa sub bab yang saling berhubungan satu dengan yang lainnya. Sistematika penulisan akan mencakup sebagai berikut:

Bab I Pendahuluan

Bab ini akan dijelaskan latar belakang, masalah yang akan diselesaikan, rumusan permasalahan, batasan tugas akhir, tujuan dan manfaat yang dihasilkan, relevansi dan sistematika penyusunan dari topik yang diambil yaitu segmentasi pelanggan.

Bab II Tinjauan Pustaka

Pada bagian tinjauan pustaka ini, akan dijelaskan mengenai referensi-referensi yang terkait dan penjelasan teori-teori yang mengenai *clustering*, algoritma K-Means, model RFM, penggalian data, pengelolaan hubungan pelanggan, Metode Elbow, Normalisasi Min-Max, SSE, AHP, CLV, *Connectivity* dan Dunn Index.

Bab III Metodologi

Bab ini akan menjelaskan mengenai tahapan-tahapan yang dilakukan yaitu identifikasi masalah, studi literatur, pengumpulan data, pra proses data, analisa RFM, proses *clustering*, uji performa hasil *clustering*, analisa hasil segmentasi, pembuatan visualisasi, penarikan kesimpulan dan saran, penyusunan laporan. Semua tahapan akan dideskripsikan dan digambarkan jadwal pengerjaan tugas akhir pada tabel waktu.

Bab IV Perancangan

Pada bab ini, akan dijelaskan tentang rancangan dari tugas akhir yang terdiri dari proses pengumpulan data, pra proses data, pemberian bobot RFM dan konstruksi visualisasi.

Bab V Implementasi

Pada bab ini menjelaskan proses implementasi berdasarkan dari rancangan yang dibuat yaitu pra proses data, proses *clustering* dan proses pembuatan visualisasi.

Bab VI Analisis dan Pembahasan

Pada bab ini akan menjelaskan hasil dan pembahasan dari proses *clustering* K-Means, uji performa *clustering*, pemberian

bobot dan perhitungan CLV. Bab ini juga berisi analisis hasil proses *clustering* dan visualisasi hasil *clustering*.

Bab VII Kesimpulan dan Saran

Bab ini memaparkan kesimpulan yang diperoleh dari rangkaian pengerjaan tugas akhir yang telah dilakukan serta menyertakan saran untuk pengembangan tugas akhir dengan topik serupa.

Halaman ini sengaja dikosongkan

BAB II TINJAUAN PUSTAKA

Pada bagian tinjauan pustaka ini, akan dijelaskan mengenai referensi-referensi yang terkait dan penjelasan teori-teori yang digunakan dalam penyusunan tugas akhir ini.

2.1. Studi Sebelumnya

Penelitian ini menggunakan penelitian terdahulu sebagai pedoman dan referensi dalam melaksanakan proses-proses dalam pengerjaan penelitian yaitu disajikan pada Tabel 2.1 hingga Tabel 2.5. Informasi yang disampaikan dalam tabel berikut berisi informasi penelitian seperti judul paper, penulis atau peneliti, tahun publikasi, deskripsi umum penelitian, dan keterkaitan terhadap penelitian tugas akhir.

Tabel 2.1 Studi sebelumnya judul paper 1

Judul Paper 1	Applying Data Mining and RFM Model to Analyze Customers' Values of A Veterinary Hospital [14]
Penulis; Tahun	Jo Ting Wei, You-Zhen Yang, Shih-Yen Lin dan Hsin-Hung Wu; 2016
Deskripsi Umum Penelitian	Tujuan: untuk mengidentifikasi pelanggan yang berharga dan mengembangkan strategi pemasaran. Studi Kasus: rumah sakit hewan di Taiwan, khususnya untuk pelanggan yang memiliki hewan anjing Metode: Variabel yang digunakan berdasarkan transaksi yang dianalisis RFM. Teknik <i>Self Organizing Maps</i> (SOM) dilakukan untuk mengetahui jumlah <i>cluster</i> terbaik, sehingga hasil SOM menjadi masukkan pada melakukan metode K-means. Lalu dilakukan klasifikasi dari hasil <i>clustering</i> dengan menyesuaikan hasil pada acuan label dari Ha-Taman dan Marcus,

	<p>sebuah matriks untuk menentukan jenis pelanggan.</p> <p>Hasil: Menjabarkan rekomendasi strategi pemasaran pada setiap hasil pengelompokan dan sedikit menjelaskan karakteristik setiap pelanggan berdasarkan label yang dibentuk.</p>
Keterkaitan Penelitian	<p>Penelitian ini digunakan untuk melihat pengelompokan terkait pelanggan yang menggunakan variabel dari hasil analisis RFM, dan melakukan <i>clustering</i> K-Means.</p>

Tabel 2.2 Studi sebelumnya judul paper 2

Judul Paper 2	Using Data Mining Techniques for Profiling Profitable Hotel Customers: an Application of RFM Analysis [15]
Penulis; Tahun	Aslihan Dursun dan Meltem Caber; 2016
Deskripsi Umum Penelitian	<p>Tujuan: untuk menyelidiki profil pelanggan yang menguntungkan atau <i>customer profiling</i></p> <p>Studi Kasus: pelanggan hotel yang terletak di Antalya, Turki.</p> <p>Metode: Segmentasi ini melibatkan karakteristik demografi pelanggan hotel. Variabel yang digunakan merupakan hasil analisis RFM, sedangkan untuk membagi kelompok pelanggan menggunakan algoritma K-Means. Teknik <i>Self Organizing Maps</i> (SOM) juga dilakukan untuk menentukan jumlah <i>cluster</i>.</p> <p>Hasil: menjelaskan secara detail setiap karakteristik dari hasil <i>cluster</i></p>
Keterkaitan Penelitian	<p>Penelitian ini digunakan untuk melihat pengelompokan terkait pelanggan yang menggunakan variabel dari hasil analisis RFM.</p>

Tabel 2.3 Studi sebelumnya judul paper 3

Judul Paper 3	Estimating Customer Future Value of Different Customer Segments Based on Adapted RFM Model in <i>Retailer</i> Banking Context [16]
Penulis; Tahun	Mahboubeh Khajvand, Mohammad Jafar Tarokh; 2011
Deskripsi Umum Penelitian	<p>Tujuan: untuk memberikan kerangka kerja memperkirakan nilai masa depan pelanggan</p> <p>Studi kasus: bank swasta di Iran, perusahaan dalam lingkup ritel perbankan.</p> <p>Metode: Data transaksi disesuaikan dengan model RFM. Metode untuk segmentasi pelanggan dengan memilih antara K-Means, Two-Step, dan X-Means yang memiliki nilai Dunn Index paling optimum. K-Means merupakan metode yang paling optimum, lalu dilakukan penghilangan data outlier dengan Anomaly Detection. Normalisasi parameter RFM dengan Min-Max dan pembobotan model RFM didapatkan dari proses AHP. Terakhir dengan model ARIMA dilakukan prediksi nilai masa depan setiap segmennya.</p> <p>Hasil: Tren nilai CLV di segmen yang berbeda selama enam musim dan grafik hasil peramalan.</p>
Keterkaitan Penelitian	Penelitian ini digunakan untuk melihat pengelompokan terkait data transaksi pelanggan yang menggunakan metode <i>clustering</i> dengan variabel dari hasil analisis RFM, serta melihat penggunaan dalam metode Dunn Index dan Normalisasi min-max.

Tabel 2.4 Studi sebelumnya judul paper 4

Judul Paper 4	Cluster Analysis Using Data Mining Approach to Develop CRM Methodology to Assess the Customer Loyalty [17]
Penulis; Tahun	Seyed Mohammad Seyed Hosseini, Anahita Maleki, Mohammad Reza Gholamian; 2010
Deskripsi Umum Penelitian	<p>Tujuan: untuk mengelompokkan loyalitas pelanggan dengan konsep B2B</p> <p>Studi kasus: SAPCO, salah satu memasok perusahaan manufaktur mobil terkemuka di Iran.</p> <p>Metode: Variabel yang digunakan merupakan hasil analisis RFML. <i>Clustering</i> dengan algoritma K-Means dilakukan untuk pengelompokkan dan penentuan k optimum oleh Davies-Bouldin Indeks. <i>Clustering</i> dilakukan dua kali yaitu dengan bobot dan tanpa bobot. Penentuan bobot berdasarkan kuesioner dan teknik <i>eigenvector</i>. Perangkingan tingkat loyalitas dilakukan dengan CLV. Evaluasi tingkat akurasi dengan R^2, Decision Tree dan metode Artificial Neural Network.</p> <p>Hasil: digambarkan pada diagram yang menunjukkan pembagian kelompok dan informasi penjualan</p>
Keterkaitan Penelitian	Penelitian ini digunakan untuk melihat pengelompokkan terkait pelanggan yang menggunakan analisis RFM dan algoritma K-Means.

Tabel 2.5 Studi sebelumnya judul paper 5

Judul Paper 5	Classifying the segmentation of customer value via RFM model and RS theory [18]
Penulis; Tahun	Ching-Hsue Cheng, You-Shyang Chen; 2009

Deskripsi Umum Penelitian	<p>Tujuan: untuk membuat model dalam pengklasifikasian pelanggan</p> <p>Studi kasus: C-company sebuah industri elektronik di Taiwan.</p> <p>Metode: penelitian ini, model RFM sebagai atribut kuantitatif untuk menjadi masukan, algoritma k-means untuk mendapatkan segmen pelanggan. Acuan penggalian aturan klasifikasinya menggunakan teori Rough Set dan LEM2.</p> <p>Hasil: model klasifikasi dan sedikit penjelasan dari data yang telah diklasifikasi</p>
Keterkaitan Penelitian	<p>Penelitian ini digunakan sebagai acuan dalam mendefinisikan variabel dan tahapan analisis RFM yang sesuai dengan PT. Bina Adidaya. Dengan penelitian ini juga menjadi acuan dalam tahapan <i>clustering</i> menggunakan algoritma K-Means.</p>

2.2. Dasar Teori

Landasan teori akan memberikan gambaran secara umum dari landasan penjabaran tugas akhir ini. Pada bagian ini menjelaskan mengenai teori-teori yang berkaitan dengan penelitian tugas akhir ini.

2.2.1. PT. Bina Adidaya Cabang Surabaya

PT. Bina Adidaya merupakan industri cat yang telah mendirikan cabang perusahaannya di Surabaya sejak Januari 1991. Perusahaan ini memiliki cabang lainnya di berbagai kota di pulau Jawa untuk melakukan distribusi produknya. Sedangkan, pabrik pembuatan produk PT. Bina Adidaya berlokasi di kota Tangerang. Komitmen yang dimiliki perusahaan ini yaitu memberikan kepuasan bagi para pelanggannya antara lain dengan cara memasok produk yang berkualitas dan layanan yang memuaskan. Cabang PT. Bina Adidaya yang berlokasi di Surabaya ini telah memiliki lebih dari 100 pelanggan yang tersebar pada kota Surabaya dan

sekitarnya. Pelanggan PT. Bina Adidaya tergantung pada jenis pemasaran yang dilakukan yaitu pemasaran retail atau pemasaran proyek. Pada pemasaran retail pelanggan perusahaan yaitu *retailer*. Kegiatan pemasaran retail ini merupakan menjual produk yang ditujukan kepada *retailer* penjual cat dan kemudian akan disalurkan kepada konsumen akhir. Jenis produk yang dijual untuk pemasaran retail ini berupa:

- Decorative adalah cat *waterbased* yang digunakan sebagai cat tembok dengan brand Pentalex A.W, Pentalex, Polaris dan VIM.
- Car Refinish adalah cat yang digunakan sebagai cat mobil dengan brand Penta Super Gloss, Penta Epoxy, Lux Epoxy, Lux Super Krypton, Penta Oto dan Lux Unoclear PU.
- Synthetic adalah cat yang dapat digunakan untuk pengecatan kayu dengan brand VIM Meni Besi, Hammertone, Optima Synthetic dan Duplex.
- Dan produk lainnya seperti Penta Aerosol, Zinc Chromate Primer, Penta Court, Lux Flinkote, Pentapoxy dan Pentaproof.

Sedangkan pemasaran proyek merupakan pemasaran yang dilakukan dengan cara kerjasama dengan pihak tertentu yang sedang melakukan tahap pembangunan. Sehingga pelanggan pada pemasaran ini bisa disebut sebagai konsumen akhir. Proyek yang dilakukan perusahaan ini seperti pengecatan bangunan-bangunan tinggi atau yang disebut *decorative*, pengecatan otomotif, peralatan kantor, kayu dan mebel rotan, drum dan produksi LPG tank, alat berat konstruksi dan lain-lain atau yang disebut sebagai *industrial paint*, pengecatan untuk ketahanan terhadap korosi atau yang disebut *protective coating*, pengecatan seluruh bagian kapal atau yang disebut *penta marine paint*, dan pengecatan atap bangunan atau yang disebut sebagai *penta roof coating*.

2.2.2. Pengelolaan Hubungan Pelanggan

Pengelolaan hubungan pelanggan merupakan strategi perusahaan untuk memahami dan mempengaruhi perilaku pelanggan melalui komunikasi yang intensif. Proses hubungan pelanggan memberikan struktur untuk menciptakan dan memelihara hubungan yang baik dengan pelanggan [19]. Di bidang teknologi informasi, pengelolaan hubungan pelanggan dapat dilihat sebagai integrasi teknologi dan proses bisnis yang digunakan untuk memenuhi kebutuhan pelanggan selama interaksi. Proses pengelolaan yang baik perlu memahami siapa pelanggannya serta apa yang mereka sukai dan tidak sukai. Hal tersebut bermaksud untuk mengantisipasi kebutuhan pelanggan dan menangani mereka secara proaktif. Pengelolaan hubungan pelanggan ini dapat mengetahui ketika pelanggan yang merasa tidak senang, sehingga perusahaan dapat melakukan sesuatu sebelum pelanggan benar-benar tidak puas dan pergi ke kompetitor [10].

Pengelolaan hubungan pelanggan dapat dilihat sebagai keuntungan oleh beberapa organisasi dan bahkan ada yang melihatnya sebagai kompetensi yang membantu organisasi untuk membedakan diri dari pesaing. Beberapa contoh manfaat potensial dari pengelolaan hubungan pelanggan adalah mengoptimalkan biaya penjualan, mengidentifikasi dan menargetkan pelanggan yang lebih baik, menurunkan biaya kampanye pemasaran, meningkatkan loyalitas pelanggan, meningkatkan retensi pelanggan, mengidentifikasi tren pelanggan dan pola konsumsi, serta membantu arus informasi ke mana pun dibutuhkan di seluruh organisasi [20]. Dalam mengelola hubungan dengan pelanggan terdapat tiga fase yaitu *acquire*, *enhance* dan *retain* [21]:

1. Mendapatkan pelanggan baru (*acquire*), pelanggan baru didapatkan dengan mempromosikan produk unggulan yang didukung oleh layanan terbaik perusahaan.
2. Meningkatkan profitabilitas dengan pelanggan yang telah ada (*enhance*), perusahaan berusaha menjalin hubungan dengan pelanggan melalui pemberian layanan dengan

kenyamanan yang lebih besar dengan biaya yang rendah, seperti penerapan *cross selling* dan *up selling*.

3. Mempertahankan pelanggan yang menguntungkan (*retain*), Berfokus pada layanan dengan tidak memberikan apa yang diinginkan pasar, tetapi apa yang diinginkan pelanggan.

2.2.3. Segmentasi Pelanggan

Segmentasi merupakan salah satu cara untuk memiliki komunikasi yang lebih dengan pelanggan. Tujuan dari segmentasi adalah untuk menyesuaikan produk, jasa, dan pesan pemasaran untuk setiap segmen [10]. Proses segmentasi menempatkan pelanggan sesuai dengan karakteristik kelompok pelanggan yang serupa. Segmentasi pelanggan merupakan langkah persiapan untuk mengklasifikasikan setiap pelanggan sesuai dengan kelompok pelanggan yang sudah ditetapkan [22]. Karakteristik pelanggan dapat direpresentasikan oleh beberapa kategori variabel yang terkait dengan pengelompokan, seperti berikut ini [10]:

- *Demographics*: Umur, jenis kelamin, besarnya keluarga, besarnya kediaman, siklus kehidupan keluarga, pemasukan, pekerjaan atau profesi, pendidikan, kepemilikan rumah, status sosial ekonomi, agama, kewarganegaraan.
- *Psychographics*: kepribadian, gaya hidup, nilai-nilai, sikap.
- *Behaviour*: manfaat yang dicari, status pembelian, tingkat penggunaan produk, frekuensi pembelian
- *Geographic*: negara, provinsi, kota, kode pos, iklim

Skema segmentasi yang berbeda dapat dikembangkan menurut tujuan bisnis yang spesifik dari organisasi. Segmentasi umumnya digunakan melalui riset data pasar untuk mendapatkan wawasan tentang sikap pelanggan, keinginan, pandangan, preferensi, dan opini tentang perusahaan dan kompetisi [8]. Segmentasi pelanggan berdasarkan pada riset pasar dan demografi seringkali membutuhkan pemahaman karakteristik semua pelanggan agar lebih efektif mengetahui

segmen apa yang menjadi menarik pelanggan. Penggalan data dapat mengembangkan segmentasi pelanggan yang juga mengidentifikasi segmentasi pada perilaku pelanggan [10]. Selain data penelitian eksternal atau pasar, data transaksi dan pembayaran pelanggan juga dapat digunakan untuk mendapatkan wawasan tentang perilaku pelanggan. Segmentasi dengan cara tersebut, dapat mengalokasikan pelanggan untuk membentuk kelompok berdasarkan jumlah pengeluaran mereka. Hal ini dapat digunakan untuk mengidentifikasi pelanggan yang bernilai tinggi dan memprioritaskan pelayanan [8].

2.2.4. Penggalan Data

Penggalan data atau yang biasa disebut dengan penggalan data merupakan proses eksplorasi dan analisis untuk menemukan pola yang berarti dan aturan pada data yang berukuran besar [10]. Data tersebut berasal dalam suatu basis data, *data warehouse*, atau tempat penyimpanan data lainnya [23]. Pengetahuan yang digunakan untuk menyajikan pengetahuan kepada pengguna. Proses penggalan data biasanya dilakukan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika. Penggalan data sebagai langkah penting dalam proses penemuan pengetahuan (*knowledge discovery process*). Urutan proses penemuan pengetahuan yaitu seperti langkah-langkah berikut [23]:

1. Pembersihan data merupakan proses awal yang dilakukan dalam penemuan pengetahuan dengan menghapus data yang dianggap mengganggu penelitian (*noise*) dan data yang tidak konsisten.
2. Integrasi data merupakan proses ini dilakukan karena dimana beberapa sumber data harus digabungkan menjadi satu.
3. Pemilihan data yaitu proses memperoleh data yang relevan pada basis data dengan memilih data yang dianggap relevan dengan analisis dalam penelitian.
4. Transformasi data yaitu proses mengubah format data sesuai keperluan dalam proses penemuan pengetahuan,

dimana data ditransformasikan dan dikonsolidasikan ke dalam bentuk yang sesuai dengan melakukan ringkasan atau agregasi operasi

5. Penggalian data merupakan proses penting dan utama karena metode cerdas akan diterapkan untuk mengekstrak pola data.
6. Evaluasi pola untuk mengidentifikasi pola yang benar-benar menarik dan mewakili pengetahuan berdasarkan ukuran *interestingness*.
7. Presentasi pengetahuan merupakan proses visualisasi dan representasi

Tugas dalam penggalian data terbagi menjadi dua kategori, yaitu penggalian data langsung (*directed*) dan penggalian data tidak langsung (*undirected*) [10]. Penggalian data langsung tujuannya yaitu menggunakan data yang tersedia untuk membangun sebuah model yang menggambarkan salah satu variabel tertentu yang menarik dalam data tersebut. Terdapat tiga tugas yang termasuk dalam penggalian data langsung yaitu *classification*, *estimation*, dan *prediction*. Penggalian data tidak langsung berarti tidak terdapat variabel yang dipilih sebagai target karena tujuannya untuk membangun beberapa hubungan pada semua variabel. Beberapa tugas yang termasuk dalam penggalian data tidak langsung adalah *clustering*, *affinity grouping* atau *association rules*, *description* dan *visualization*. Pada penelitian ini proses penggalian data digunakan untuk membantu pengolahan data transaksi pelanggan PT. Bina Adidaya. Tugas penggalian data pada penelitian ini termasuk pada penggalian data tidak langsung yaitu *clustering* karena bertujuan untuk membuat kelompok dengan menghubungkan semua variabel pada data yang ada.

2.2.5. *Clustering*

Clustering merupakan proses untuk mengelompokkan sekumpulan objek data yang memiliki kemiripan tinggi ke dalam satu kelompok atau *cluster* dan setiap kelompok yang terbentuk tidak memiliki kemiripan dengan kelompok lainnya [23]. Kemiripan dinilai berdasarkan nilai atribut yang

mendeskripsikan objek data. *Clustering* atau juga disebut sebagai segmentasi ini adalah salah satu metode penggalian data yang *unsupervised*, karena tidak ada atribut yang digunakan sebagai panduan atau tidak adanya label pada data dalam proses pembelajaran. Teknik dalam *clustering* dilakukan untuk menemukan pengetahuan dari kumpulan data [24]. Analisis *cluster* dalam konteks penggalian data yaitu proses menempatkan pelanggan atau prospek kedalam kelompok yang memiliki ciri-ciri yang sama [10].

Algoritma *clustering* membangun sebuah model dengan melakukan serangkaian pengulangan dan berhenti ketika model tersebut telah terpusat dan batasan dari segmentasi telah stabil. Hasil dari *clustering* yang bagus tergantung dengan ukuran kesamaan dan metode yang digunakan. Pendekatan dalam *cluster* berdasarkan saran dari Fraley dan Raftery, membagi metode pengelompokan menjadi dua kelompok utama yaitu metode hirarkis dan metode partisi [25].

1. Metode Hirarki yaitu metode yang membentuk *cluster* dengan mempartisi secara berulang-ulang dari atas ke bawah atau sebaliknya. Hasil dari metode hirarki berupa dendogram yang mewakili kelompok objek dan tingkat kesamaan di mana terdapat perubahan pengelompokan. Sebuah pengelompokan objek data diperoleh dengan memotong dendrogram pada tingkat kemiripan yang diinginkan [26].
2. Metode Partisi yaitu metode yang membuat inisial partisi k di awal, dimana parameter k merupakan jumlah partisi untuk membentuk. Kemudian secara iterative menggunakan teknik relokasi dengan mencoba berulang-ulang memindahkan objek dari satu kelompok ke kelompok lain untuk mendapatkan partisi yang optimal. Tipe metode partisi ini seperti K-Means, K-Medoids dan CLARANS [23].

Sedangkan menurut Kamber terdapat tambahan tiga metode mengenai cluster yaitu *Density Based Method*, *Model Based* dan *Grilde Based* [23].

1. *Density Based Method* yaitu metode yang mengasumsikan bahwa titik milik setiap cluster diambil secara spesifik dari distribusi probabilitas. Tujuan metode ini adalah untuk mengidentifikasi kelompok dan distribusi parameternya [27]
2. *Model Based* yaitu metode yang mengoptimalkan kesesuaian antara data dan beberapa model matematika serta mencari deskripsi karakteristik setiap kelompok dimana setiap kelompok mewakili sebuah *cluster*. Metode yang sering digunakan *Decision Trees* dan *Neural Networks* [26].
3. *Grice Based* yaitu mempartisi ruang menjadi sel-sel dengan jumlah yang terbatas dan nantinya akan membentuk struktur jaringan dimana semua oprasi untuk *clustering* dilakukan. Metode memiliki keuntungan yaitu waktu memproses yang singkat [23].

Pada penelitian ini proses merode *clustering* yang digunakan yaitu metode partisi karena bertujuan untuk membuat kelompok yang setiap pelanggannya hanya berada pada satu kelompok tertentu.

2.2.6. Algoritma K-Means

Algoritma K-means merupakan salah satu algoritma *clustering* yang bertujuan untuk membagi data menjadi beberapa kelompok. Algoritma ini menerima masukan berupa data tanpa label kelas (*unsupervised learning*). Proses *clustering* K-Means, dilakukan oleh komputer dengan mengelompokan sendiri data-data yang menjadi masukannya tanpa mengetahui terlebih dulu target kelasnya. Pada setiap *cluster* terdapat titik pusat (centroid) yang merepresentasikan *cluster* tersebut. Algoritma untuk melakukan K-Means *clustering* adalah sebagai berikut [12]:

1. Menentukan nilai k sebagai jumlah *cluster* yang dibentuk
2. Menentukan nilai awal centroid atau titik pusat *cluster*. Pada tahapan ini ditentukan nilai centroid secara acak, namun untuk tahapan berikutnya menggunakan rumus 2.1.

$$\overline{V}_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \quad (2.1) [28]$$

Dimana:

V_{ij} = cantroid *cluster* ke-i untuk variabel ke-j

N_i = banyaknya data pada cluster ke-i

i,k = indeks dari *cluster*

j = indeks dari variabel

X_{kj} = nilai data ke-k yang ada di dalam *cluster* untuk variabel ke-j

3. Menghitung jarak antara titik centroid dengan titik setiap objek, dapat dilakukan dengan menggunakan Euclidean Distance dengan rumus 2.2.

$$D_e = \sqrt{(x_i - s_i)^2 + (y_t - t_t)^2} \quad (2.2) [28]$$

Dimana:

De = Euclidean Distance

i = banyaknya data

(x,y) = koordinat data

(s,t) = koordinat centroid.

4. Mengelompokan data hingga terbentuk *cluster* dengan titik centroid dari setiap *cluster* merupakan titik centroid yang terdekat. Penentuan anggota *cluster* adalah dengan memperhitungkan jarak minimum objek.
5. Memperbarui nilai centroid setiap *cluster*.
6. Mengulangi langkah ke-2 hingga akhir sampai nilai dari titik centroid tidak lagi berubah.

2.2.7. Metode Elbow

Metode Elbow memberikan ide atau gagasan dengan cara memilih nilai *cluster* dan kemudian menambah nilai *cluster* tersebut untuk dijadikan model data dalam penentuan *cluster* terbaik. Metode ini akan menghasilkan informasi dalam menentukan jumlah *cluster* terbaik dengan cara melihat persentase hasil perbandingan antara jumlah *cluster* yang

membentuk siku pada suatu titik. Algoritma Metode Elbow dalam menentukan nilai k pada K-Means [29].

1. Inisialisasi \$ k = 1 \$
2. Mulai
3. kenaikan nilai \$ k \$
4. Mengukur biaya solusi kualitas optimal
5. Jika di beberapa titik biaya solusi turun drastis
6. Hal tersebut menunjukan nilai \$ k \$ yang benar
7. Selesai

Hasil persentase yang berbeda dari setiap nilai *cluster* dapat ditunjukkan dengan menggunakan grafik sebagai sumber informasinya. Grafik akan menunjukkan beberapa nilai k yang mengalami penurunan paling besar dan selanjutnya hasil dari nilai k akan turun secara perlahan-lahan sampai hasil dari nilai k tersebut stabil. Namun kelemahan dari metode ini yaitu tidak selalu dapat teridentifikasi titik sikunya [13].

2.2.8. Normalisasi Min-Max

Metode Min-Max merupakan metode normalisasi sederhana dengan melakukan transformasi terhadap data asli. Min-Mix akan menyesuaikan batas yang ditentukan dengan menghubungkan pada data asli [30]. Teknik normalisasi ini melakukan transformasi sebuah atribut numerik dalam range atau skala yang lebih kecil seperti 0.0 sampai 1.0 [31]. Angka tersebut menunjukkan bahwa batas terendah 0.0 sedangkan batas tertinggi 1.0. Normalisasi Min-Max dapat dihitung dengan persamaan 2.3.

$$x' = \frac{x - \min_a}{\max_a - \min_a} (\text{newmax} - \text{newmin}) + \text{newmin} \quad (2.3) \quad [32]$$

Dimana:

- x' = nilai yang telah dinormalisasi
- x = nilai mentah yang akan dinormalisasikan
- \min_a = nilai terendah pada setiap variabel
- \max_a = nilai tertinggi pada setiap variabel
- newmax = rentang maksimal X yang bernilai 1
- newmin = rentang minimal X yang bernilai 0

Kelebihan dari Min-Max yaitu nilai perbandingan antar data sebelum normalisasi seimbang dengan data sesudah normalisasi dan kelebihan lainnya tidak ada data bias yang dihasilkan. Kelemahan dari Min-Max, jika ada data baru, akan memungkinkan terjebak pada "out of bound" error [33].

2.2.9. Sum of Square Error (SSE)

SSE merupakan metode yang sederhana dan banyak digunakan dalam mengukur kriteria pada evaluasi pengelompokan karena SSE mengukur keseragaman antar data dalam satu *cluster*. Metode ini menggunakan informasi yang ada pada objek saja, sehingga termasuk ke dalam kriteria kualitas internal. Semakin kecil nilai SSE maka semakin bagus hasil *clustering*-nya. Rumus untuk mencari SSE terdapat pada persamaan 2.4.

$$SSE = \sum_{i=1}^K \sum_{x \in C_i} dist^2(m_i, x) \quad (2.4) [12]$$

Dimana:

k = banyak *cluster* yang terbentuk

C_i = *cluster* ke- i

m_i = *cluster* ke- i

x = data yang ada pada setiap *cluster*

2.2.10. Dunn Index

Dunn Index merupakan metrik untuk validitas hasil kelompok yang ditentukan dengan menggunakan diameter kelompok dan jarak antara dua kelompok. Metode ini biasanya digunakan untuk mengidentifikasi kekompakan dan pemisahan *cluster*, sehingga hasil dari *cluster* dapat dihitung kembali. Perhitungan *Dunn Index* dilakukan dengan mengukur rasio jarak terkecil antar kelompok dengan jarak terbesar di dalam kelompok. Perhitungan *Dunn Index* dilakukan dengan persamaan 2.5.

$$DU_k = \min_{i=1, \dots, k} \left\{ \min_{j=i+1, \dots, k} \left(\frac{diss(c_i, c_j)}{\max_{m=1, \dots, k} (diam(c_m))} \right) \right\} \quad (2.5) [34]$$

Dimana:

DU = *Dunn Index*

k = jumlah kelompok

- $diss(c_i, c_j)$ = jarak kelompok i dan kelompok j yang dihitung pada persamaan
- $diam(c_m)$ = diameter kelompok i yang dihitung pada persamaan

Dunn Indeks akan memeriksa kesesuaian *cluster*, varian antar *cluster*, dan memberikan nilai untuk pengelompokan tertentu. Jumlah *cluster* yang optimal memiliki nilai Dunn Index yang tertinggi [35]

2.2.11. Connectivity

Nilai yang didapatkan dari metode *Connectivity* menunjukkan kekuatan interaksi antar anggota klaster. Konektivitas memiliki nilai antara nol dan ∞ dan jumlah *cluster* yang optimal didapatkan pada nilai *connectivity* yang paling minimal atau rendah. Indeks konektivitas didefinisikan sebagai 2.6.

$$Conn = \sum_{i=1}^N \sum_{j=1}^L x_{i,nn_{i(j)}} \quad (2.6) [36]$$

Dimana:

$nn_{i(j)}$ = tetangga terdekat dari j dari pengamatan di i.

$x_{i,nn_{i(j)}}$ = akan mendekati nol jika i dan j berada di cluster yang sama dan $1/j$ sebaliknya.

L = parameter yang menentukan jumlah tetangga yang berkontribusi terhadap indeks konektivitas.

2.2.12. Model RFM

Analisa model ini diperkenalkan pertama kali oleh Hughes dan saat ini banyak digunakan oleh industri termasuk manufaktur, *retailer*, dan industri jasa [37]. RFM ini bertujuan untuk menentukan segmentasi pelanggan berdasarkan tiga variabel yaitu *Recency of the last purchases*, *Frequency of the purchases*, dan *Monetary value of the purchases* [38].

- *Recency* merupakan jarak antara waktu terakhir transaksi dengan waktu saat ini. Apabila jarak semakin kecil maka nilai R semakin besar.

- *Frequency* merupakan seberapa sering jumlah transaksi yang dilakukan oleh pelanggan pada periode tertentu, misalnya seperti tiga kali dalam periode satu tahun. Apabila jumlah transaksi makin besar maka nilai F juga makin besar.
- *Monetary* berarti jumlah uang yang dihabiskan pelanggan saat transaksi pada periode tertentu, jika jumlah uang makin besar maka nilai M juga makin besar.

Semakin besar nilai R dan F maka kemungkinan pelanggan yang akan melakukan transaksi kembali dengan perusahaan tersebut. Selain itu semakin besar nilai M, maka kecenderungan pelanggan dalam memberikan respon kepada produk dan layanan perusahaan tersebut [9]. Perusahaan menghitung skor RFM setiap pelanggan untuk menentukan kemungkinan pelanggan akan merespon dengan baik misalnya mengenai penawaran, promosi atau katalog. Banyak perusahaan percaya bahwa pelanggan yang telah menjadi pembeli baru dan paling sering serta telah menghabiskan sebagian besar dalam jangka waktu tertentu yang paling mungkin untuk merespon positif penawaran perusahaan di masa depan [39]. Skor tersebut dapat menentukan misalnya apakah seorang pelanggan akan dikirimkan katalog mahal atau hanya berupa kartu pos.

2.2.12.1. Pembobotan Model RMF

Pada pemberian bobot ini terdapat dua macam analisa yaitu pertama ketika variabel RFM memiliki kepentingan yang sama sehingga bobot ketiganya bernilai sama, sedangkan kedua ketika variabel RFM memiliki tingkat kepentingan yang berbeda-beda atau bobot yang berbeda tergantung dengan karakteristik dari perusahaan [18]. Pembobotan yang berbeda pada setiap variabelnya juga dapat ditentukan dengan berbagai cara yaitu seperti berdasarkan pengukuran yang didapatkan dari pengalaman sehingga mengetahui kepentingan setiap variabel dan ada juga yang diambil berdasarkan proses *Analytic Hierarchy Process* (AHP) [40].

2.2.13 Pembobotan Menggunakan *Analytical Hierarchy Process* (AHP)

AHP adalah suatu proses untuk mengembangkan nilai numerik untuk peringkat setiap alternatif keputusan, berdasarkan seberapa baik setiap alternatif memenuhi kriteria pembuat keputusan. Metode peringkat alternatif keputusan ini akan memilih yang terbaik ketika pembuat keputusan memiliki beberapa tujuan, atau kriteria, yang menjadi dasar dalam keputusan [41]. Langkah yang dilakukan untuk mendapatkan bobot dengan metode AHP yaitu [42]:

1. Membuat *pairwise comparison* dengan melakukan survei terlebih dahulu melalui kuisioner untuk memberikan nilai setiap kriteria yang dipasangkan mengacu pada *preference scale* di
2. Tabel 2.6.

Tabel 2.6 Preference Scale

Skala Nilai	Tingkat Preferensi	Penjelasan
1	Sama pentingnya	Kedua kriteria memiliki pengaruh yang sama
3	Sedikit lebih penting	Penilaian dan pengalaman sedikit memihak pada salah satu kriteria tertentu dibandingkan kriteria pasangannya.
5	Lebih penting	Penilaian dan pengalaman memihak pada salah satu kriteria tertentu dibandingkan kriteria pasangannya.
7	Jelas lebih penting	Salah satu kriteria lebih diprioritaskan dan relatif lebih penting dibandingkan kriteria pasangannya.
9	Mutlak sangat penting	Salah satu kriteria sangat jelas lebih penting dibandingkan kriteria pasangannya.
2, 4, 6, 8		Diberikan bila terdapat keraguan penilaian di antara dua tingkat kepentingan yang berdekatan.

3. Setelah menentukan prioritas masing-masing kriteria menggunakan matriks *pairwise comparison*, kemudian melakukan normalisasi matriks. Pertama-tama menambahkan jumlah setiap kolom. Kemudian membagi setiap elemen dengan hasil penjualan perkolomnya. Matriks baru yang dihasilkan disebut normal comparisons matrix.
4. Menghitung bobot setiap kriteria dengan cara menghitung rata-rata angka setiap baris dalam normal comparisons matrix.
5. Melakukan uji rasio konsistensi yaitu instrumen yang menentukan konsistensi dan menunjukkan tingkat kepercayaan prioritas yang diperoleh dari perbandingan. Terdapat dua tahapan dalam uji rasio konsistensi [41]:
 - a. Setiap elemen pada tabel yang berisi matriks *pairwise comparison* dikalikan dengan bobot dari setiap kriteria yang didapatkan pada langkah ke-3, lalu hitung *overall score* dengan menjumlahkan setiap barisnya.
 - b. Mencari nilai x sebagai masukan untuk menghitung CI dengan cara mengalikan *overall score* dengan bobot kemudian hasilnya dirata-rata.
 - c. Menghitung indeks konsistensinya dengan persamaan 2.7.

$$CI = \frac{x-n}{n-1} \quad (2.7) [41]$$

Dimana:

CI = indeks konsistensi

x = hasil dari langkah 4.b

n = banyaknya kriteria

Apabila telah konsisten maka nilai indeks adalah 0. Apabila nilai indeks bukan 0 maka perlu melakukan proses selanjutnya

- d. Menghitung tingkat ketidakkonsistenan yang dapat diterima dengan persamaan 2.8.

$$\frac{CI}{RI} \quad (2.8) [41]$$

Dimana:

CI = indeks konsistensi

RI = random indeks

Tabel 2.7 Random Indeks

n	2	3	4	5	6	7
RI	0	0.58	0.90	1.12	1.24	1.32
	n	8	9	10		
	RI	1.41	1.45	1.51		

Tingkat ketidakkonsistenan diterima apabila nilai yang didapatkan dari perhitungan CI/RI lebih kecil dari 0.1.

2.2.14. Customer Lifetime Value (CLV)

Customer Lifetime Value (CLV) atau yang sering dikenal sebagai *Life Time Value* (LTV) merupakan pemahaman mengenai nilai kekinian dari seluruh keuntungan masa depan yang diperoleh dari pelanggan. *Customer Lifetime Value* harus dihitung pada tingkat segmen pelanggan yang disesuaikan dengan proses bisnis. [43]

Penerapan *customer lifetime value* dalam tugas akhir ini, pendekatan yang digunakan yaitu indeks nilai CLV. Metodologi yang dilakukan yaitu dengan menggunakan RFM terbobot berdasarkan penilaian dari level manager pada perusahaan terkait yang melalui proses analisis hierarki (AHP), maka nilai indeks CLV didapatkan dari persamaan 2.9.

$$CLV_{ci} = NR_{ci} \times WR_{ci} + NF_{ci} \times WF_{ci} + NM_{ci} \times WM_{ci} \quad (2.9)$$

Dimana:

$NR_{ci}, NF_{ci}, NM_{ci}$ = secara berturut-turut menggambarkan nilai normalisasi *recency*, *frequency*, dan *monetary*.

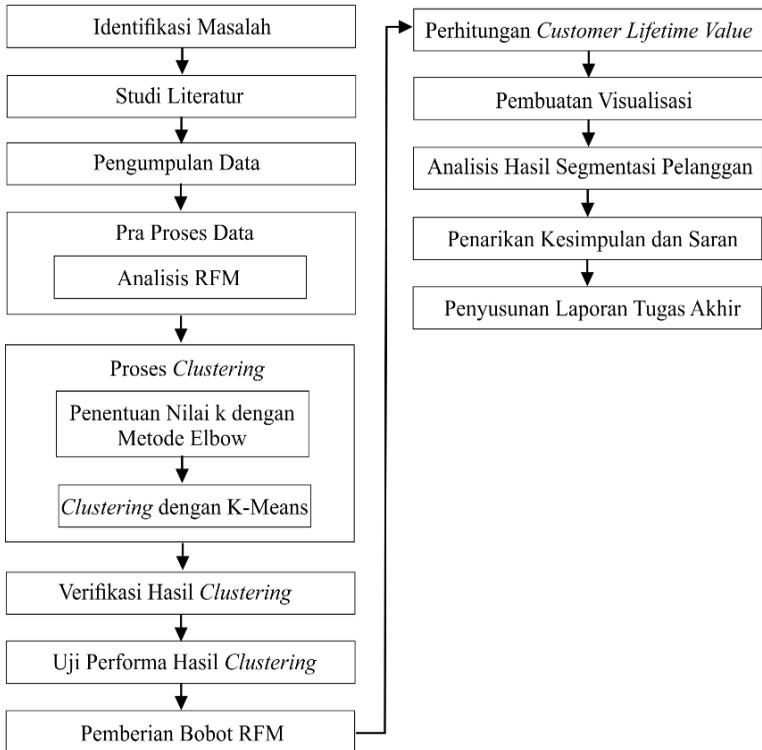
$WR_{ci}, WF_{ci}, WM_{ci}$ = secara berurutan merupakan bobot dari nilai *recency*, *frequency*, dan *monetary*.

BAB III METODOLOGI PENELITIAN

Pada bab ini akan menjelaskan tentang metodologi yang akan digunakan dengan deskripsi setiap tahapannya.

3.1. Metodologi

Metodologi digunakan sebagai panduan dalam penyusunan tugas akhir agar terarah dan sistematis. Adapun urutan dari pengerjaan tugas akhir pada Gambar 3.1 berikut ini.



Gambar 3.1 Metodologi tugas akhir

3.1.1 Identifikasi Masalah

Tahapan ini merupakan tahapan awal penyusunan tugas akhir yaitu melakukan analisis permasalahan yang akan dijadikan topik pengerjaan tugas akhir. Objek yang dipilih dan menjadi masukan pada proses identifikasi adalah PT. Bina Adidaya cabang Surabaya. Keluaran dari proses identifikasi berupa permasalahan yang diangkat yaitu segmentasi pelanggan untuk masukan dalam merencanakan strategi pengelolaan hubungan pelanggan khususnya mempertahankan *retailer*.

3.1.2 Studi Literatur

Studi literatur didapatkan dari pengumpulan referensi, diantaranya dari narasumber, buku, *e-book*, penelitian sebelumnya seperti pada jurnal dan paper, tugas akhir, artikel dan dokumen yang terkait dengan segmentasi pelanggan. Segmentasi ini akan melakukan proses penggalian data, sehingga referensi yang diperlukan berhubungan dengan *clustering*, algoritma K-Means, model RFM, penggalian data, pengelolaan hubungan pelanggan, Metode Elbow, Normalisasi Min-Max, SSE, AHP, CLV, *Connectivity* dan *Dunn Index*. Keluaran yang didapatkan pada proses studi literatur berupa kajian pustaka sebagai acuan dalam konsep, metode, landasan teori untuk pembelajaran pengerjaan dalam tugas akhir.

3.1.3 Pengumpulan Data

Pada tahapan ini melakukan pengumpulan data-data yang dibutuhkan sebagai pendukung utama dalam pengerjaan tugas akhir. Data yang diambil disesuaikan dengan topik dan batasan permasalahan pada tugas akhir. Dalam hal ini, data didapatkan dari PT. Bina Adidaya cabang Surabaya diantaranya adalah:

- Laporan pengeluaran barang berdasarkan surat jalan bulan Januari 2016 hingga Desember 2016
- Laporan berdasarkan surat jalan dan *invoice* pada bulan Januari 2016 hingga Desember 2016
- Data *retailer* PT. Bina Adidaya di wilayah Surabaya

Selain pengumpulan data, dilakukan juga wawancara kepada narasumber terkait kondisi terkini perusahaan yang berhubungan dengan segmentasi sebagai informasi pendukung dalam penelitian tugas akhir.

3.1.4 Pra Proses Data

Pada tahap ini akan dilakukan pemrosesan data mentah untuk menyesuaikan kebutuhan. Pra proses data ini akan melakukan pemilihan atribut, membersihkan baris data yang memiliki nilai kosong, menggabungkan data mentah yang diperoleh, selain itu terdapat proses analisis RFM. Keluaran pada proses ini berupa data yang siap untuk *clustering*.

3.1.4.1. Analisis RFM

Pada saat melakukan transformasi terdapat beberapa tahap diantaranya mengubah nilai kedalam bentuk RFM. Mencari nilai atribut *recency*, *frequency* dan *monetary* dengan cara melakukan agregasi operasi menggunakan query pada alat bantu aplikasi Navicat. Hasilnya yang diperoleh akan diubah format file dan melakukan normalisasi data hasil pencarian nilai RFM.

3.1.5 Proses Clustering

Pada proses clustering dilakukan dengan alat bantu aplikasi RStudio. Hasil dari analisa RFM yang telah dinormalisasi akan diolah terlebih dahulu menggunakan metode Elbow sebagai proses menentukan nilai k kemudian melakukan *clustering* dengan K-Means. Keluaran *cluster* yang terbentuk akan diberi label atau nama untuk memudahkan perusahaan dalam mengingat karakteristik pelanggannya.

3.1.9.1. Penentuan Nilai k dengan Metode Elbow

Tahap pertama bertujuan untuk membantu penentuan nilai k sebagai masukan dalam penerapan K-Means. Metode yang digunakan yaitu Metode Elbow. Proses awal yaitu melakukan inisiasi rentang nilai k yang akan diolah dalam metode ini. Sehingga keluaran yang didapatkan berupa nilai k yang dipilih dari hasil grafik yang menunjukkan titik siku.

3.1.9.2. Clustering dengan K-Means

Proses *clustering* K-Means dilakukan untuk menemukan segmen pelanggan yang dimiliki oleh PT. Bina Adidaya. Masukan proses K-Means ini merupakan hasil analisis RFM yang telah dinormalisasi dan hasil nilai k yang didapatkan pada metode Elbow. Lalu, penentuan nilai centroid dan jarak setiap objek data ke centroid. Sehingga keluarannya berupa *cluster* dengan memiliki centroid yang telah tetap atau tidak berpindah.

3.1.6 Verifikasi Hasil Clustering

Proses yang dilakukan untuk memastikan bahwa setiap *retailer* yang dikelompokkan tepat berada dalam segmen tersebut. Tahapan ini dilakukan dengan menghitung jarak *retailer* ke titik pusat kelompok dengan menggunakan algoritma *Euclidean Distance*.

3.1.7 Uji Performa Hasil Clustering

Pada tahap ini hasil *cluster* akan mengukur tingkat performa model *cluster* yang telah terbentuk pada tahap sebelumnya. Pada uji performa ini akan mengetahui seberapa baik *cluster* dipisahkan dan seberapa dekat objek yang berhubungan dalam satu *cluster* uji prforma ini dilakukan secara internal yaitu dengan menggunakan perhitungan SSE, *Dunn Index* dan *Connectivity*.

3.1.8 Pemberian Bobot RFM

Pembobotan ketiga atribut di penelitian ini dilakukan melalui proses AHP. Sebelumnya masukan dari AHP ini merupakan hasil survey berupa kuisioner yang diisikan oleh pihak perusahaan. Hasil dari AHP akan memberikan bobot yang berbeda-beda pada variabel *recency*, *frequency* dan *monetary*. Hasil pembobotan akan dilakukan uji rasio konsistensi untuk mengetahui konsistensi dari bobot yang telah dihitung.

3.1.9 Perhitungan Customer Lifetime Value (CLV)

Tahap selanjutnya yaitu menghitung indeks CLV, namun terlebih dahulu mencari rata-rata nilai normalisasi dari *recency*,

frequency dan *monetary* pada setiap *cluster*. Nilai yang didapatkan dikalikan dengan bobot yang dihasilkan dari metode AHP. Indeks CLV didapatkan dari menjumlahkan nilai variabel RFM pada setiap *cluster*. Besar kecilnya indeks CLV besar menentukan tingkat loyalitas dari pelanggan.

3.1.10 Pembuatan Visualisasi

Tahapan ini mencakup pembuatan *interface* yang berbasis *web-based* dengan alat bantu RShiny. Visualisasi ini bertujuan untuk membaca hasil *clustering* sehingga dapat merencanakan pengelolaan hubungan pelanggan menjadi lebih cepat. Keluaran visualisasi ini disajikan dengan diagram dan keterangan pada setiap kelompok yang terbentuk. Keterangan yang diberikan dapat berupa rentang indeks dan karakteristik setiap segmen.

3.1.11 Analisis Hasil Segmentasi Pelanggan

Setiap *cluster* yang terbentuk akan dianalisis untuk mengetahui karakteristik *retailer*. Hasil analisis akan menjelaskan ciri khas setiap segmennya berupa kesamaan perilaku pelanggan PT. Bina Adidaya dan secara tidak langsung akan menjelaskan perbedaan dengan segmen yang lainnya. Selain itu, juga dilakukan analisa untuk hasil visualisasi grafik dan gambar yang dibuat.

3.1.12 Penarikan kesimpulan dan Saran

Setelah visualisasi berbasis *web-based* telah dibuat, maka dibuatlah saran-saran dari hasil penelitian untuk PT. Bina Adidaya dan pengembangan untuk penelitian selanjutnya agar memberikan luaran yang jauh lebih baik.

3.1.13 Penyusunan Laporan Tugas Akhir

Tahapan terakhir adalah penyusunan laporan tugas akhir sebagai bentuk dokumentasi bukti telah selesainya penelitian tugas akhir ini. Seluruh proses pengerjaan yang dilakukan di dalam penelitian ini akan didokumentasikan dalam sebuah buku dengan format mengikuti peraturan yang berlaku di Departemen Sistem Informasi ITS.

Halaman ini sengaja dikosongkan

BAB IV PERANCANGAN

Pada bab ini, akan dijelaskan tentang rancangan dari tugas akhir yang terdiri dari proses pengumpulan data, pra proses data, pemberian bobot RFM dan konstruksi visualisasi.

4.1. Pengumpulan Data

Pengumpulan data dilakukan setelah melalui proses diskusi dan persetujuan dari kepala cabang PT. Bina Adidaya Surabaya. Proses pengumpulan data dilakukan dengan wawancara secara langsung terhadap salah satu karyawan pada divisi I untuk menggali proses bisnis dari perusahaan terkait pemasaran *retail* dan kendala dalam proses pemasaran yang selama ini dihadapi. Proses berikutnya yaitu pengambilan data yang dilakukan oleh karyawan di divisi I. Data tersebut merupakan data transaksi *retailer* pada periode 7 Januari 2016 hingga 30 Desember 2016. Data diperoleh berbentuk rekapan excel sebanyak tiga dokumen yaitu:

- Data *retailer* PT. Bina Adidaya di wilayah Surabaya dengan 337 baris data
- Laporan pengeluaran barang berdasarkan surat jalan dengan 6950 baris
- Laporan berdasarkan surat jalan dan *invoice* dengan 1860 baris data.

Total atribut dari ketiga dokumen berbeda-beda. Jika secara keseluruhan dijumlahkan, banyak atribut yang ada pada semua data yaitu sebanyak 20 atribut Tabel 4.1 merupakan potongan data *retailer* pelanggan PT. Bina Adidaya, Tabel 4.2 merupakan potongan laporan pengeluaran barang berdasarkan surat jalan dan Tabel 4.3 merupakan laporan berdasarkan surat jalan dan *invoice*. Data transaksi yang masih mentah ini diberikan dalam file excel. Penjelasan mengenai atributnya terdapat pada Lampiran B.

Tabel 4.1 Potongan data *retailer* pelanggan PT. Bina Adidaya area Surabaya

Code	Debtor Name	Address1	Address2	City
GA1.0001	A.A #	Jl. Simo Jawar 53	Telp: 7492141, 7492142	Surabaya
GA1.0002	Anugerah	Jl. Sukomanunggal 150	Telp: 7492162, 7497767, 7496975	Surabaya
GA1.0003	Bayu Santoso	Jl. Simorejo Sari B - 64	Telp:	Surabaya
GA1.0004	Berkat	Jl. Raya Benowo 97	Telp. 7420524	Surabaya
GA1.0005	Cahaya Sari	Jl. Kedung Doro 235	Telp:	Surabaya

Tabel 4.2 Potongan laporan pengeluaran barang berdasarkan surat jalan

Packing	SJ No.	SJ Update	Customer	Qty
1 LTR	GA1-06-0107/008	01/08/2016	Gunung Rejeki	12
			Total ----->	12
5 LTR	GA1-06-0107/008	01/08/2016	Gunung Rejeki	12
5 LTR	GA1-06-0112/002	01/12/2016	Sumber Rejeki	30
			Total ----->	42
0,640 LTR	GA1-06-0120/002	20/01/16	Sumber Rejeki II	72
0,640 LTR	GA1-06-0120/002	20/01/16	Sumber Rejeki II	72
			Total ----->	144
1 KG	GA1-06-0122/004	23/01/16	Gallery Cat / Cv. Cahaya Bintang Sejati	12

Tabel 4.3 Potongan laporan berdasarkan surat jalan dan *invoice*

No.	SJ No. Product Code	DO No. Customer Product Name	Pack. Size (Unit)	Quantity		Price Per Unit (Rp)	Total (Rp)
				ltr	kg		
1	GA1-06- 0107/001	DO-601/ 06003-O Wisma Harapan Nc Black	1 kg		60	70.400,00	4.224.000,00
	5460 .03.0020	Nc Black Flat	1 kg		36	70.400,00	2.534.400,00
	5860 .03.0020M	Nc Black	0,2 kg		12	218.350,00	1.091.750,00
	5460 .27.0020	Nc Black Flat	0,2 kg		12	218.350,00	1.091.750,00
	5860 .27.0020M	Nc Dark Chrome Yellow	0,2 kg		4,8	218.350,00	436.700,00
2	GA1-06- 0107/002	DO-601/ 05004-O Wisma Harapan Hammertone Alkyd Black	1 ltr	36		57.200,00	2.059.200,00
3	GA1-06- 0107/003	DO-601 /05005-O Cv. Seng Joyo Mandiri Zinc Chromate Prim.Blue 307	5 kg		100	159.940,00	3.198.800,00
4	GA1-06- 0107/004	Do- 601/06002- O Cv. Seng Joyo Mandiri Nc Black	1 kg		36	70.400,00	2.534.400,00
	5460 .03.0020						

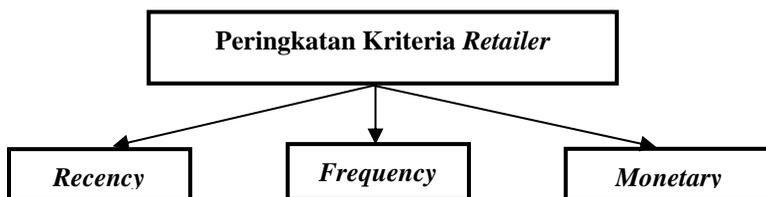
4.2. Pra Proses Data

Data mentah yang diperoleh dari PT. Bina Adidaya akan dilakukan penyesuaian yang mengacu pada langkah awal dalam proses penggalan data. Pra proses data ini meliputi:

1. Pemilihan data yaitu untuk menyesuaikan atribut yang digunakan dengan model RFM. Atribut yang digunakan seperti tanggal *update* surat jalan dan jumlah biaya yang dikeluarkan pelanggan. Penambahan lainnya diluar model RFM seperti data identitas pelanggan dan nama pelanggan.
2. Pembersihan data untuk menghilangkan baris data yang memiliki nilai kosong atau bisa disebut data yang tidak valid.
3. Integrasi data untuk menggabungkan antara data laporan pengeluaran barang dengan data dari laporan berdasarkan surat jalan dan *invoice* serta data *retailer*.
4. Transformasi data dalam penelitian ini terdapat tiga proses.
 - a. Proses analisis RFM untuk mengubah atribut data sesuai dengan model RFM pada setiap *retailer*. Nilai atribut *recency* didapat dari menghitung kapan transaksi terakhir pada data tanggal. Nilai *frequency* dari menghitung banyaknya transaksi pada data tanggal. Nilai *monetary* didapat dari menjumlahkan biaya secara keseluruhan.
 - b. Proses mengubah format file data dari excel yaitu *xlsx* menjadi *csv* untuk keperluan dalam memasukan data set di aplikasi Rstudio.
 - c. Proses normalisasi data untuk mengatasi skala perbedaan yang cukup jauh pada setiap variabel khususnya antara variabel *monetary* dengan variabel *recency* atau *frequency*. Metode normalisasi dalam penelitian yaitu Min-Max yang dihitung pada setiap variabel. Proses normalisasi dilakukan di Rstudio. Setelah normalisasi, untuk variabel *recency* nilainya dibalik agar memiliki makna yang sama dengan variabel *frequency* dan *monetary*.

4.3. Pemberian Bobot RFM

Proses analisis RFM selanjutnya berkaitan dengan proses awal untuk menganalisis CLV yaitu pemberian bobot pada nilai RFM menggunakan AHP. Rancangan dalam pemilihan kriteria disusun seperti Gambar 4.1. Data dalam proses AHP diperoleh dari pengisian kuisioner. Responden yang mengisi kuisioner ada empat orang yaitu kepala cabang, *supervisor*, kepala admin dan kepala gudang PT. Bina Adidaya Surabaya. Rancangan kuisioner berisi data responden (nama, jabatan dan lama bekerja), petunjuk pengisian kuisioner beserta dengan penjelasan RFM dan *preference scale* yang mengacu pada bab II, contoh pengisian dan tabel perbandingan kriteria dengan skor yang ada pada *preference scale*. Hasil pembuatan isi dari kuisioner dapat dilihat pada Lampiran D.



Gambar 4.1 Kriteria pemilihan

4.4. Konstruksi Visualisasi

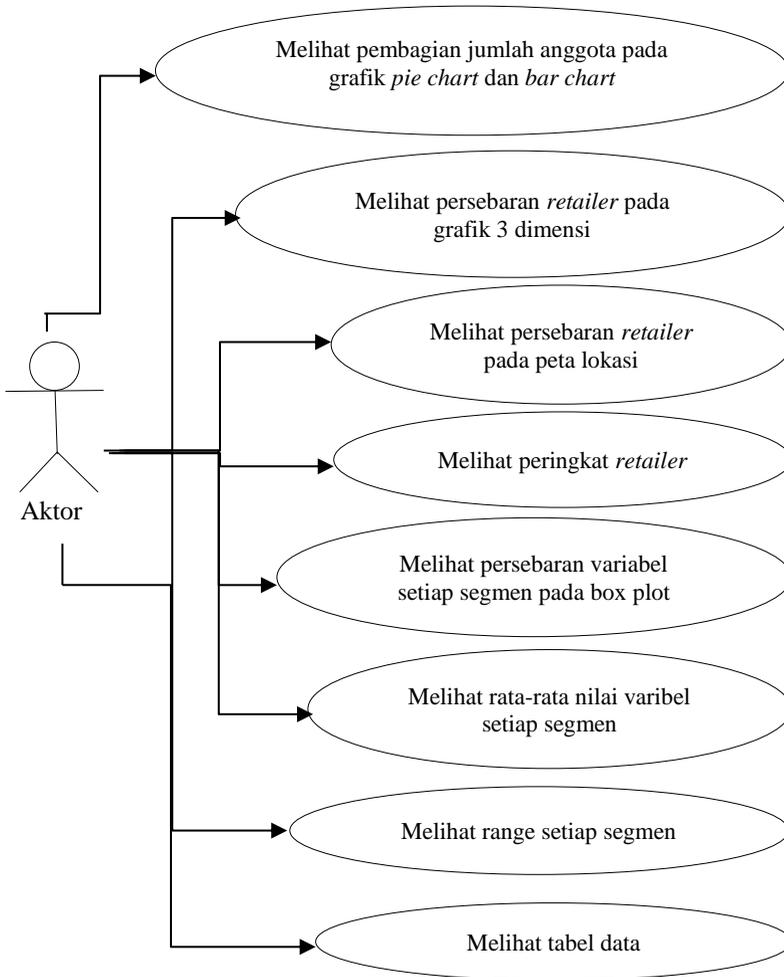
Perancangan visualisasi hasil segmentasi *retailer* PT. Bina Adidaya Surabaya meliputi pembuatan use case diagram dan use case description.

4.4.1. Use Case Diagram

Visualisasi yang dibuat untuk menjelaskan setiap diagram yang dibuat dan informasi karakteristik dari setiap *retailer*, sehingga Gambar 4.2 merupakan use case diagram untuk perancangan visualisasi yang terdapat 8 use case utama, antara lain:

- Melihat pembagian jumlah anggota pada grafik *pie chart* dan *bar chart*
- Melihat persebaran *retailer* pada grafik 3 dimensi
- Melihat persebaran *retailer* pada peta lokasi

- d. Melihat peringkat *retailer*
- e. Melihat persebaran variabel setiap segmen pada box plot
- f. Melihat range setiap segmen
- g. Melihat rata-rata nilai variabel setiap segmen
- h. Melihat tabel data



Gambar 4.2 Use case diagram

4.4.2. Deskripsi Use Case

Deskripsi mengenai use case melihat pembagian jumlah anggota pada grafik *pie chart* dan *bar chart* dalam konstruksi visualisasi akan dijelaskan lebih rinci pada Tabel 4.4 sedangkan tujuh use case utama lainnya dijelaskan pada Lampiran E

Tabel 4.4 Use case description “Melihat pembagian jumlah anggota pada grafik *pie chart* dan *bar chart*”

UC-1 Melihat Pembagian Jumlah Anggota pada grafik <i>pie chart</i> dan <i>bar chart</i>		
Purpose	Mengetahui jumlah anggota setiap segmen melalui grafik <i>pie chart</i> dan <i>bar chart</i>	
Overview	Dimulai ketika data hasil proses <i>clustering</i> telah dimasukkan pada script aplikasi	
Actors	Pengguna aplikasi	
Pre Condition	Pengguna telah membuka aplikasi	
Past Condition	Grafik “Jumlah Anggota Segmen” dan “Persentase Jumlah Anggota Segmen”	
Typical Course Event	Actor	System
	1. Klik icon slide bar  3. Klik menu 2 Segmen atau 3 Segmen	2. Menampilkan slide bar pilihan menu 2 Segmen atau 3 Segmen 4. Menghasilkan grafik Jumlah Anggota Segmen dan Persentase Jumlah Anggota Segmen
Alternate Flow of Events	Step 2: tidak mengklik menu Segmen 2 atau Segmen 3, maka aplikasi akan memilih menu default yaitu segmen 2	
Exceptional Flow of Event	Aplikasi tertutup otomatis karena error	

4.4.3. Desain Antarmuka Visualisasi

Penjelasan mengenai desain antarmuka visualiasasi akan dijelaskan perhalaman. Visualisasi ini terdiri dari 6 halaman namun berisi konten yang sama hanya berbeda pada data yang dimasukkan yaitu ketika di *cluster* menjadi 2 dan data ketika di *cluster* menjadi 3, sehingga halaman yang berisi konten berbeda yaitu halaman segmantasi pelanggan, halaman detail segmen dan halaman tabel data.

A. Halaman 2 Segmen dan 3 Segmen

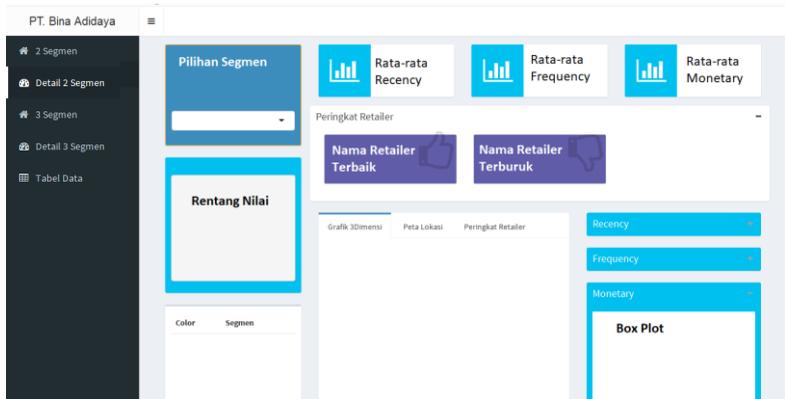
Gambar 4.3 merupakan halaman untuk melihat pembagian jumlah *retailer* pada setiap segmennya dalam bentuk diagram batang dan *pie chart*. Selain itu pada halaman ini, memberikan informasi mengenai banyak *retailer* yang diproses, batasan yang digunakan seperti area pelanggan, data transaksi yang digunakan dan kategori produk.



Gambar 4.3 Desain antarmuka halaman 2 dan 3 segmen

B. Halaman Detail 2 Segmen dan 3 Segmen

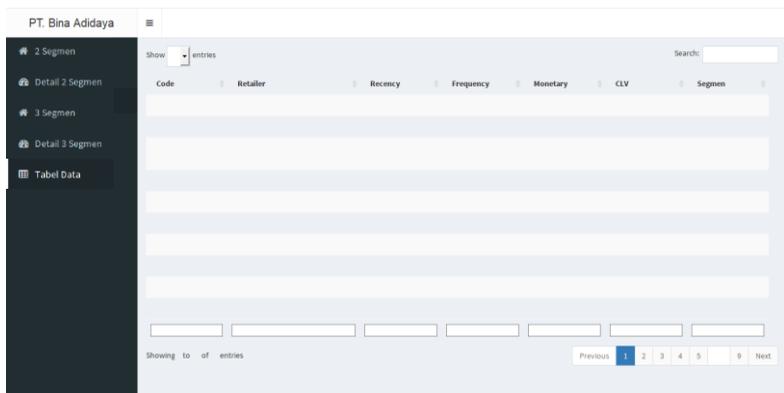
Gambar 4.4 merupakan halaman untuk melihat detail setiap segmennya. Halaman ini terdapat pilihan untuk jenis segmen yang ingin ditampilkan, terdapat tab menu yang berisi grafik 3 dimensi, gambar peta lokasi, dan tabel data nilai CLV. Terdapat juga box plot setiap variabel serta informasi mengenai range dan rata-rata variabel, selain itu juga menampilkan keterangan warna yang digunakan pada grafik dan peta.



Gambar 4.4 Desain antarmuka halaman detail 2 dan 3 segmen

C. Halaman Tabel Data

Gambar 4.5 merupakan halaman untuk melihat data secara keseluruhan dalam bentuk tabel, terdapat bantuan pencarian setiap kolom, halaman tabel dan pilihan tampilan jumlah baris.



Gambar 4.5 Desain antarmuka halaman tabel data

Halaman ini sengaja dikosongkan

BAB V IMPLEMENTASI

Pada bab ini menjelaskan proses implementasi berdasarkan dari rancangan yang dibuat yaitu pra proses data, proses *clustering* dan proses pembuatan visualisasi.

5.1. Pra Proses Data

Berdasarkan metodologi pengerjaan tugas akhir pada bab 3, tahapan untuk pra proses data terdiri dari empat aktivitas yaitu pemilihan data, integrasi data, pembersihan data dan transformasi data. Penjelasan yang lebih rinci akan dijelaskan pada setiap aktivitasnya.

5.1.1. Pemilihan Data

Pemilihan atribut pada data mentah ini dilakukan untuk menyeleksi atribut yang diperlukan dalam proses *clustering*. Sebanyak 20 atribut yang berasal dari ketiga data dipilih secara manual. Atribut yang digunakan untuk proses selanjutnya adalah SJ No, SJ Update, Code, Customer atau Debtor Name dan Total (Rp). Pada atribut Total (Rp) dilakukan agregasi penjumlahan berdasarkan nomor surat jalan di excel, sehingga menghasilkan atribut Biaya. Penjelasan mengenai fungsi dari atribut yang telah dipilih terdapat pada Tabel 5.1.

5.1.2. Pembersihan Data

Proses pembersihan data untuk menghilangkan data yang redundan dan hanya dilakukan pada data laporan pengeluaran barang berdasarkan surat jalan. Penghapusan baris ini mengacu pada kode unik dari nomor surat jalan. Penghapusan baris menggunakan *remove duplicates* pada excel, sehingga dari 6950 baris data menjadi 1860 baris data.

Tabel 5.1 Pemilihan atribut

Atribut yang memiliki nilai unik sebagai atribut untuk proses penggabungan data	SJ -laporan pengeluaran barang berdasarkan surat jalan
	Retailer -laporan pengeluaran barang berdasarkan surat jalan
	Retailer -data retailer
	SJ -laporan berdasarkan surat jalan dan invoice
	Retailer -laporan berdasarkan surat jalan dan invoice
Informasi penting dari data yang tidak didapatkan dari data lainnya	Tanggal -laporan pengeluaran barang berdasarkan surat jalan
	Code -data retailer
	Biaya -laporan berdasarkan surat jalan dan invoice

Pembersihan data juga dilakukan untuk menghilangkan baris data transaksi yang tidak memiliki nilai. Penghapusan baris ini dilakukan hanya pada laporan berdasarkan surat jalan dan *invoice* dengan cara *sort* dan *filter* pada excel yaitu mengurutkan biaya dari terkecil hingga terbesar. Setelah mengurutkan akan terlihat baris data yang memiliki nilai kosong yaitu pada laporan berdasarkan surat jalan dan *invoice*. Hasil dari pembersihan data ini yaitu sejumlah dua baris data yang memiliki biaya 0 terlihat pada Tabel 5.2, sehingga tersisa 1858 baris data.

Tabel 5.2 Baris data yang memiliki nilai biaya nol

SJ	Retailer	Biaya
GA1-0P-0903/001	CV. Sejahtera	
GA3-06-0206/002	Proline	

5.1.3. Integrasi Data

Proses awal pada integrasi data yaitu terlebih dahulu memasukkan file excel ke dalam sebuah aplikasi basis data yaitu Navicat. Integrasi ini dilakukan menggunakan query

dengan fungsi INNER JOIN yang berarti menggabungkan dan menampilkan hasil dari kedua tabel yang sama saja. Proses integrasi akan dilakukan sebanyak dua kali.

Integrasi yang pertama yaitu antara laporan pengeluaran barang berdasarkan surat jalan dengan laporan berdasarkan surat jalan dan *invoice*. Tujuan dari integrasi pertama yaitu dapat melihat ditanggal berapa dan berapa jumlah biaya yang dihabiskan *retailer* pada setiap nomor surat jalan. Atribut yang digunakan untuk integrasi adalah SJ, kemudian ditampilkan secara urut menurut SJ dengan fungsi ORDER BY. Secara lebih rinci Script 5.1 merupakan query yang digunakan.

```
SELECT Tanggal, barang.SJ,barang.Retailer, biaya
FROM barang
INNER JOIN invoice ON barang.SJ=invoice.SJ
ORDER BY barang.SJ;
```

Script 5.1 Integrasi pertama

Setelah proses penggabungan yang pertama berhasil dilakukan, maka didapatkan 1858 baris data transaksi. Banyaknya data yang terhapus juga karena pada laporan berdasarkan surat jalan dan *invoice* ada 2 baris yang memiliki nilai biaya kosong.

Integrasi berikutnya yaitu antara hasil integrasi pertama dengan data *retailer*. Atribut yang digunakan untuk penggabungan adalah nama *retailer*, kemudian ditampilkan secara urut menurut SJ dengan fungsi ORDER BY. Secara lebih rinci Script 5.2 merupakan query yang digunakan.

```
SELECT Tanggal, SJ, retailer.`Code`,retailer
.Retailer, biaya
FROM joinawal
INNER JOIN retailer ON joinawal.Retailer=
retailer.Retailer
ORDER BY SJ;
```

Script 5.2 Integrasi kedua

Setelah proses dua kali penggabungan berhasil dilakukan, maka didapatkan 1667 baris data transaksi. Banyaknya data yang terhapus juga karena masih terdapat transaksi dari *retailer* yang bukan area Surabaya.

5.1.4. Transformasi Data

Dalam transformasi ini terdapat empat proses yaitu melakukan analisis RFM, normalisasi dan pengubahan format file. Penjelasan yang lebih rinci akan dijelaskan pada setiap aktivitasnya.

5.1.4.1. Analisis RFM

Analisis RFM merupakan proses untuk mencari nilai atribut yaitu *recency*, *frequency* dan *monetary*. Pencarian nilai atribut melakukan agregasi menggunakan query dan dijalankan pada aplikasi Navicat.

Nilai R merupakan selisih antara waktu saat pengerjaan yaitu 20 Maret 2017 dengan waktu terakhir melakukan transaksi. Atribut yang dibutuhkan yaitu tanggal dan code. Nilai R didapatkan dari tanggal yang paling terakhir menggunakan fungsi MAX berdasarkan jenis *retailer* dengan fungsi GROUP BY, sedangkan untuk mengetahui tanggal saat ini menggunakan fungsi NOW dan mencari selisihnya menggunakan fungsi DATEDIFF.

Nilai F merupakan nilai yang menggambarkan berapa kali jumlah transaksi pelanggan. Atribut yang dibutuhkan yaitu tanggal dan code. Nilai F didapatkan dari menghitung banyaknya tanggal transaksi dengan fungsi COUNT berdasarkan jenis *retailer* dengan fungsi GROUP BY.

Nilai M merupakan total biaya yang dikeluarkan *retailer* untuk melakukan transaksi. Atribut yang dibutuhkan yaitu biaya dan code. Nilai M didapatkan dari menjumlahkan biaya dengan fungsi SUM berdasarkan jenis *retailer* dengan fungsi GROUP BY.

Nilai R, F, dan M ditampilkan secara urut menurut code *retailer* dengan fungsi ORDER BY. Secara lebih rinci berikut merupakan query yang digunakan untuk mencari nilai RFM untuk setiap *retailer*.

```

SELECT
`Code`, retailer, DATEDIFF(NOW(), MAX(Tanggal)) AS
R, COUNT(DISTINCT Tanggal) AS F, SUM(Biaya) AS M
FROM hasiljoin
GROUP BY `Code`
ORDER BY `Code`;

```

Script 5.3 perubahan nilai ke dalam bentuk RFM

Setelah mengeksekusi query dan berhasil mendapatkan nilai atribut RFM pada setiap *retailer*, jumlah baris data yang di dapatkan sebanyak 83 baris. Data tersebut menunjukkan bahwa terdapat 83 jenis *retailer* dari total *retailer* 337 yang melakukan transaksi pada periode Januari 2016 hingga Desember 2016.

5.1.4.2. Perubahan Format File

Pengubahan format file ini dilakukan untuk kebutuhan dalam memasukkan data set ke dalam aplikasi Rstudio yaitu proses *x clustering* dan visualisasi. Format awal yang masih berupa Excel 97-2003 Workbook (*.xls) diubah menjadi format Comma Delimited (*.csv). Proses perubahan dilakukan dengan fungsi save as pada Microsoft Excel.

5.1.4.3. Normalisasi Data

Normalisasi dilakukan agar selisih nilai antar pelanggan tidak terlalu jauh. Proses normalisasi ini dilakukan di aplikasi Rstudio. Metode normalisasi Min Max mentransformasikan seluruh nilai atribut menjadi rentang nilai antara 0 hingga 1. Normalisasi ini akan dilakukan pada setiap variabel R, F dan M dengan menggunakan nilai minimal dan maksimal setiap variabel. Nilai maksimal dan minimal dari setiap atribut tercantum pada Tabel 5.3.

Algoritma perhitungan yang dilakukan mengacu rumus 2.3 di bab II, nilai setiap baris dikurangi dengan nilai minimal pada setiap kolom lalu hasilnya dibagi dengan selisih antara nilai maksimal dan minimal setiap kolom. Secara lebih rinci berikut merupakan Script 5.4 yang dijalankan untuk normalisasi ini.

Tabel 5.3 Nilai maksimal dan minimal setiap variabel

Maksimum Recency	436
Minimum Recency	80
Maksimum Frequency	119
Minimum Frequency	1
Maksimum Monetary	910.232.412
Minimum Monetari	422.730

```

> rnorm <- (adidaya$recency-min(adidaya$recency))/
(max(adidaya$recency)-min(adidaya$recency))
> fnorm <- (adidaya $frequency-min(adidaya$frequen
cy))/(max(adidaya $frequency)-min(adidaya $frequen
cy))
> mnorm <- (adidaya$monetary-min(adidaya$monetary)
)/(max(adidaya$monetary)-min(adidaya$monetary))
> hasil <- data.frame(adidaya,rnorm,fnorm,mnorm)
> write.csv(hasil,"D:/LILY/adidaya.csv")

```

Script 5.4 Normalisasi Min-Max

Nilai dari *recency* memiliki makna terbalik dengan *frequency* dan *monetary*. Nilai *recency* yang kecil merupakan *recency* yang terbaik sedangkan nilai *frequency* dan *monetary* terkecil merupakan nilai terburuk. Oleh karena itu, ketiga nilai dalam variabel tersebut disamakan terlebih dahulu dengan membalik nilai *recency* setiap *retailer*. Proses membalik ini dengan cara 1 dikurangi dengan nilai *recency* setiap *retailer*. Hasil akhir dari pra proses secara keseluruhan terdapat pada Lampiran A.

5.2. Proses *Clustering*

Dalam proses *clustering* terdapat dua proses didalamnya yaitu menentukan nilai k dengan metode Elbow dan melakukan *clustering* dengan metode K-Means. Kedua proses tersebut dilakukan pada aplikasi Rstudio. Sebelum melakukan kedua proses tersebut perlu memasukkan data set numerik dan melakukan inisiasi nama data set untuk memudahkan dalam

proses perhitungan. Proses tersebut dijalankan menggunakan Script 5.5.

```
> adidaya <- read.csv("D:/TA/ adidaya.csv", sep=";")
> data <- data.frame(adidaya$rba1ik, adidaya$fnorm, ad
idaya$mnorm)
```

Script 5.5 Input dan inisiasi data set

5.2.1. Penentuan Nilai k dengan Metode Elbow

Data set yang telah dilakukan pra proses data merupakan data yang telah siap untuk digunakan, selanjutnya dilakukan pemilihan jumlah *cluster* atau nilai k dengan menggunakan metode Elbow. Metode ini memilih jumlah *cluster* dengan melihat penurunan secara signifikan pada nilai SSE dan di titik itu nilai SSE sudah mulai stabil (tidak turun terlalu signifikan). Titik tersebut yang menjadi titik siku pada grafik.

Proses dari metode Elbow ini menggunakan percobaan jumlah *cluster* antara 1 sampai 10. Proses terakhir melihat hasil plot untuk mengetahui titik siku yang terbentuk. Rincian proses tersebut dijalankan menggunakan Script 5.6.

```
> elbow <- sapply(1:10, function(k){kmeans(data, k, ns
tart=10 )$tot.withinss})
> plot(1:10, elbow,type="b", pch = 19, xlab="Number of
clusters k",ylab="Total within-clusters sum of squares
")
```

Script 5.6 Metode Elbow

5.2.2. Clustering dengan K-Means

Proses *clustering* dilakukan setelah mendapatkan nilai k. Proses ini dimulai dengan membuat inisiasi nama, inisiasi ini berisi fungsi untuk menjalankan K-Means pada RStudio dan memasukkan nilai k sebanyak 3. Hasilnya proses K-Means dapat ditampilkan dengan memanggil inisiasi tersebut. Untuk mendapatkan hasil yang lebih mudah dipahami, maka dilakukan penggabungan antara data dan hasil *clustering*

menggunakan fungsi `data.frame` dan disimpan kedalam bentuk file dengan format csv menggunakan fungsi `write.csv`. Rincian proses tersebut dijalankan menggunakan Script 5.7. Proses pengelompokkan menjadi 2 segmen sama dengan Script 5.7 hanya saja nilai `k` 3 diganti dengan nilai `k` 2.

```
> klaster<-kmeans(data,3)
> klaster
> output<-data.frame(adidaya,klaster$cluster)
> write.csv(output, "hasil3klaster.csv")
```

Script 5.7 *Clustering* 3 segmen menggunakan K-Means

5.3. Verifikasi Hasil Clustering

Proses untuk mengetahui bahwa setiap *retailer* tepat berada pada segmen yang dibentuk, maka perlu dilakukan verifikasi data menggunakan rumus *Euclidean Distance*. Setiap *retailer* akan dihitung jaraknya ke centroid. Perhitungan jarak dengan *Euclidean Distance* dapat mengetahui seberapa dekat *retailer* ke setiap centroid.

Script 5.8 merupakan verifikasi ini dilakukan pada aplikasi R untuk 3 segmen. Proses ini membutuhkan 3 titik centroid dan titik setiap *retailer*. Perbedaan proses verifikasi untuk 2 segmen dan 3 segmen yaitu pada 2 segmen hanya menggunakan centroid sebanyak dua, sehingga proses perhitungan jarak setiap datanya hanya ke dua centroid.

Tabel centroid dipisahkan menjadi tabel centroid per*cluster*. Kolom baru ditambahkan untuk menghitung jarak setiap *retailer* ke setiap titik centroid. Proses mencari nilai yang minimal dalam kolom dan barisnya yaitu mengambil tiga kolom atau jumlah kolom yang sesuai dengan banyaknya centroid yang berisi hasil perhitungan jarak dan menampilkan kolom ke berapa dengan nilai yang minimal. Hasilnya digabungkan dengan data hasil *clustering* dan menyesuaikan bahwa kolom yang minimal sama dengan hasil *clustering*.

```

#mengambil setiap baris centroid
View(datacentroid)
CK1<-datacentroid[1,]
CK2<-datacentroid[2,]
CK3<-datacentroid[3,]

#cari distance
Datak1 <- hasilsegmen
Datak1$distk1 <-sqrt(
  ((hasilsegmen$rbalik-CK1$recency)^2)
  +((hasilsegmen$fnorm-CK1$frequency)^2)
  +((hasilsegmen$mnorm-CK1$monetary)^2)
)
Datak1$distk2 <-sqrt(
  ((hasilsegmen$rbalik-CK2$recency)^2)
  +((hasilsegmen$fnorm-CK2$frequency)^2)
  +((hasilsegmen$mnorm-CK2$monetary)^2)
)
Datak1$distk3 <-sqrt(
  ((hasilsegmen$rbalik-CK3$recency)^2)
  +((hasilsegmen$fnorm-CK3$frequency)^2)
  +((hasilsegmen$mnorm-CK3$monetary)^2)
)
View(Datak1)

#memilih yang minimal pake 3 kolom aja
Eudist<-data.frame(Datak1$distk1,Datak1$distk2,Datak1$distk3)
Eudist$hasilcek<-apply(Eudist,1,which.min) #cari kolom yg min
Eudist$min<-apply(Eudist,1,min)
View(Eudist)

#gabung
hasilcek <- Eudist
hasilcek$klasterke <- hasilsegmen$klasterke
hasilcek$code <- hasilsegmen$code
hasilcek$retailer <- hasilsegmen$retailer
View(hasilcek)

write.csv(hasilcek,"D:/REVISI/ceksegmen/hasilcek.csv")

```

Script 5.8 Euclidean Distance 3 segmen

5.4. Uji Performa *Clustering*

Data hasil *clustering* dengan K-Mean yang pada mulanya membagi ke dalam 3 segmen perlu diuji peformanya untuk mengetahui apakah 3 segmen tersebut telah optimal menggunakan metode lainnya. Metode yang diterapkan dalam pengujian performa yaitu metode SSE, *Connectivity* dan *Dunn Index*.

Pada metode SSE, data yang diperlukan hanya nilai setiap atribut R,F, dan M yang telah dinormalisasi. Algoritma SSE akan menghitung nilai kuadrat dari jarak setiap data ke centroid lalu menghitung total secara keseluruhan. Hasil dari metode SSE ini akan ditampilkan dalam plot diagram dan nilai SSE

setiap *cluster*. Hasil perhitungan dapat disimpan dalam bentuk file dengan format csv. Proses perhitungan SSE menggunakan Script 5.9.

```
> data <- data.frame(adidaya$rbalik, adidaya$fnorm,
  adidaya$mnorm)
> wss = kmeans(data, centers=1)$tot.withinss
> for (i in 2:10)wss[i] = kmeans(data, centers=i)$tot
  .withinss
> plot(wss,xlab="Number of clusters",ylab="within-clu
  sters sum of squares")
> write.csv(wss,"sse.csv")c
```

Script 5.9 Metode Sum of Square Error

Pada metode *Dunn Index* dan *Connectivity*, data yang diperlukan hanya nilai setiap atribut R, F, dan M yang telah dinormalisasi serta kolom yang berisikan nama seriap baris. Metode *Dunn Index* dan *Connectivity* menggunakan package *clValid*. Metode *Dunn Index* akan memilih jumlah *cluster* yang memiliki nilai paling maksimal. Sedangkan untuk metode *Connectivity* akan memilih jumlah *cluster* yang memiliki nilai paling minimal. Proses perhitungan tersebut dijalankan menggunakan Script 5.10.

```
> data <- data.frame(adidaya$rbalik, adidaya$fnorm, ad
  idaya$mnorm)
> rownames(data) <- adidaya$code
> library(clValid)
> x <- clValid(data, 2:10, clMethods="kmeans", validat
  ion="internal")
> summary(x)
> plot(x, legend=FALSE)
```

Script 5.10 Metode Conectivity dan Dunn Index

5.5. Pemberian Bobot RFM

Proses pemberian bobot ini membutuhkan data yang telah didapatkan dari pengisian kuisisioner. Data tersebut di rekap dengan cara merata-rata semua pengisian responden dengan

melihat perbandingan setiap kriterianya, sehingga hasilnya seperti pada Tabel 5.4.

Tabel 5.4 *Pairwise comparisons matrix*

Kriteria	Recency	Frequency	Monetary
Recency	1	1/4	2 3/4
Frequency	4 1/4	1	7 1/2
Monetary	2/5	1/7	1
Column Sum	5 2/3	1 3/8	11 1/4

Proses selanjutnya yaitu melakukan normalisasi matriks. Normalisasi dilakukan dengan menjumlahkan setiap kolom, lalu membagi setiap elemen dengan hasil penjualan perkolomnya. Matriks baru yang dihasilkan disebut *normal comparisons matrix*, dan hasil bobot yang didapatkan setiap kriteria dihitung dengan cara merata-rata kriteria disetiap barisnya seperti pada Tabel 5.5.

Bobot yang didapatkan perlu dilakukan uji rasio konsistensi untuk mengetahui apakah pengisian kuisioner oleh responden telah bernilai konsisten atau tidak. Setiap elemen pada tabel *pairwise comparisons matrix* dikalikan dengan bobot dari setiap kriteria. *overall score* dihitung dengan menjumlahkan setiap barisnya seperti yang ditunjukkan pada

Tabel 5.6.

Tabel 5.5 *Normal comparison matrix* dan bobot

<i>Normal Comparisons Matrix</i>				Bobot
Kriteria	Recency	Frequency	Monetary	
Recency	0.18	0.18	0.24	0.20
Frequency	0.75	0.72	0.67	0.71
Monetary	0.07	0.10	0.09	0.09

Tabel 5.6 Perhitungan *overall score*

Kriteria	Recency	Frequency	Monetary	Overall Score
Recency	0.20	0.18	0.23	0.614
Frequency	0.85	0.71	0.64	2.207
Monetary	0.08	0.10	0.09	0.260

Dari hasil perhitungan *overall score* dapat dicari nilai x sebagai masukan untuk menghitung CI. Caranya yaitu mengalikan setiap baris dari *overall score* dengan bobot kemudian hasilnya dirata-rata.

$$\text{Recency} = 0.614 / 0.20 = \mathbf{3.057}$$

$$\text{Frequency} = 2.207 / 0.71 = \mathbf{3.090}$$

$$\text{Monetary} = 0.260 / 0.09 = \mathbf{3.054}$$

$$\text{Average atau nilai } x = (3.057 + 3.090 + 3.054) / 3 = \mathbf{3.067}$$

Setelah mengetahui nilai x yang menjadi masukan untuk menghitung CI, maka dengan mengacu rumus 2.7 pada bab II, sehingga kriteria tersebut memiliki CI sebesar:

$$CI_{kriteria} = \frac{3.067 - 3}{3 - 1} = 0.034$$

Didapatkan CI sebesar 0.034 dan nilai tersebut menunjukkan CI yang didapatkan tidak sama dengan 0. Oleh karena itu, perlu menghitung tingkat ketidakkonsistensi yang dapat diterima. Caranya yaitu mengacu rumus pada 2.8 di bab II. Jumlah kriteria ada tiga, maka pada tabel random indeks $n=3$ bernilai 0.58, sehingga kriteria tersebut memiliki CI/RI sebesar 0.058.

$$\frac{0.034}{0.58} = 0.058$$

5.6. Perhitungan CLV

Perhitungan CLV membutuhkan data yang telah di *cluster* dan hasil *cluster* telah diuji performanya. Proses perhitungan ini perlu mencari nilai rata-rata indeks setiap variabel. Nilai tersebut didapatkan dari menghitung rata-rata nilai RFM pada setiap *cluster* atau nilai centroid setiap *cluster*. Tabel 5.8 merupakan tabel perhitungan rata-rata atau centroid untuk 2 segmen sedangkan Tabel 5.8 merupakan tabel rata-rata atau centroid untuk 3 segmen.

Tabel 5.7 Nilai rata-rata RFM atau centroid 2 segmen

<i>Cluster</i>	CR	CF	CM
1	0.3296731	0.01155624	0.009282096
2	0.9007184	0.14656849	0.115991921

Tabel 5.8 Nilai rata-rata RFM atau centroid 3 segmen

<i>Cluster</i>	CR	CF	CM
1	0.954002809	0.492584746	0.643374498
2	0.887796504	0.092749529	0.035781998
3	0.31540931	0.011702986	0.009547113

Proses selanjutnya membutuhkan bobot yang telah didapatkan dari perhitungan AHP. Indeks CLV dihitung dengan menjumlahkan antara hasil perkalian rata-rata nilai indeks RFM yang sudah dihitung sebelumnya dengan bobot RFM. Nilai indeks CLV 2 segmen ditunjukkan dalam Tabel 5.9 sedangkan nilai indeks CLV 3 segmen ditunjukkan dalam

Tabel 5.10.

Tabel 5.9 Peringkat CLV 2 segmen

<i>Cluster</i>	CR x WR	CF x WF	CM x WM	Indeks CLV	Peringkat CLV
1	0.06593	0.00820	0.00074	0.0749	2
2	0.18014	0.10406	0.00928	0.29349	1

Tabel 5.10 Peringkat CLV 3 segmen

<i>Cluster</i>	CR x WR	CF x WF	CM x WM	Indeks CLV	Peringkat CLV
1	0.19080	0.34974	0.05147	0.59201	1
2	0.17756	0.06585	0.00286	0.24627	2
3	0.06308	0.00831	0.00076	0.07215	3

5.7. Pembuatan Visualisasi dengan RShiny

Proses visualisasi dilakukan menggunakan aplikasi Rshiny. *Package* yang perlu dijalankan dalam aplikasi yaitu *package shiny* dan untuk desainnya menggunakan *package shiny-dashboard*. Penjelasan lebih rinci mengenai isi dari setiap halaman akan dijelaskan pada setiap aktivitas berikut.

5.7.1. Halaman 2 Segmen dan 3 Segmen

Pada halaman 2 segmen dan 3 segmen sama-sama terdapat dua grafik yaitu *bar chart* dan *pie chart*, untuk penjelasan lebih rinci mengenai pembuatan setiap grafik akan dijelaskan pada setiap aktivitas.

5.6.1.1. Visualisasi dengan Bar Chart

Proses pembuatan *bar chart* membutuhkan *package* *plotly* dan data yang digunakan yaitu data hasil *clustering*. Sumbu x akan menunjukkan 3 jenis segmen yang terbentuk. Sumbu y menunjukkan proses hasil perhitungan banyaknya baris yang ada di setiap jenis *cluster*. Tampilan *bar chart* diberi warna yang sama dengan grafik lainnya yaitu kuning, biru dan merah. Untuk membuat visualisasi dengan *bar chart*, maka Script 5.11 yang dijalankan pada halaman 3 segmen.

```
#Data Bar dan Pie
segmen <- c("Segmen 1","Segmen 2","Segmen 3")
value <- c(sum(ta.data$klasterke == "1"),
           sum(ta.data$klasterke == "2"),
           sum(ta.data$klasterke == "3"))

#warna Cluster
warna <- c("#D2691E", "#1E90FF", "#DC143C")

#BAR
output$bar <- renderPlotly ({
  plot_ly(
    x = segmen,
    y = value,
    name = "Jenis Segmen",
    type = "bar",
    hoverinfo = 'text',
    text = paste(value, ' Retail'),
    marker = list(color = warna)
  )
})
```

Script 5.11 Visualisasi dengan Bar Chart 3 segmen

5.6.1.2. Visualisasi dengan Pie Chart

Proses pembuatan *pie chart* membutuhkan *package* *plotly* dan data yang digunakan yaitu data hasil *clustering*. Angka yang didapatkan sama dengan sumbu y di *bar chart* yaitu banyaknya baris yang ada di setiap jenis *cluster* dalam satuan persentase. Tampilan *pie chart* diberi warna yang sama dengan *bar chart*. Untuk membuat visualisasi dengan *pie chart*, maka Script 5.12 yang dijalankan pada halaman 3 segmen.

```
#Data Bar dan Pie
segmen <- c("Segmen 1", "Segmen 2", "Segmen 3")
value <- c(sum(ta.data$klasterke == "1"),
           sum(ta.data$klasterke == "2"),
           sum(ta.data$klasterke == "3"))
#Warna Cluster
warna <- c("#D2691E", "#1E90FF", "#DC143C")

#PIE
output$pie <- renderPlotly ({
  plot_ly(labels = segmen, values = value, type = 'pie',
          textposition = 'inside',
          textinfo = 'label+percent',
          insidetextfont = list(color = '#FFFFFF'),
          hoverinfo = 'text',
          text = paste(value, ' Retail'),
          marker = list(colors = warna,
                       line = list(color = '#FFFFFF', width = 1))
  })
})
```

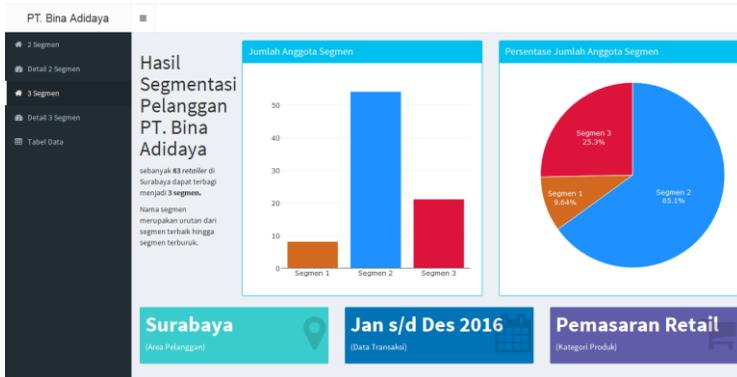
Script 5.12 Visualisasi dengan *pie chart* 3 segmen

Gambar 5.1 merupakan hasil tampilan visualisasi halaman 2 segmen berdasarkan pada desain antarmuka yang dibuat sebelumnya.



Gambar 5.1 Halaman 2 segmen

Gambar 5.2 merupakan hasil tampilan visualisasi halaman 3 segmen berdasarkan pada desain antarmuka yang dibuat sebelumnya.



Gambar 5.2 Halaman 3 segmen

5.7.2. Halaman Detail 2 dan 3 Segmen

Pada halaman detail 2 segmen dan detail 3 segmen sama-sama terdapat dua grafik, sebuah gambar peta, sebuah tabel dan informasi. Grafik tersebut berupa grafik 3 dimensi dan box plot, untuk penjelasan lebih rinci mengenai pembuatan setiap isi yang ada pada halaman ini akan dijelaskan pada setiap aktivitas.

5.6.2.1. Visualisasi dengan Grafik 3 Dimensi

Proses pembuatan grafik ini menggunakan plot 3 dimensi yaitu plot3D, sehingga membutuhkan *package* plot3Drgl. Data yang digunakan yaitu data hasil *clustering*, nilai *recency*, nilai *frequency* dan nilai *monetary* masing-masing *retailer*. Tampilan grafik untuk sumbu x berupa nilai *recency*, sumbu y berupa nilai *frequency* dan sumbu z berupa nilai *monetary*. Setiap titik berupa bola kecil menunjukkan letak *retailer* berdasarkan nilai RFM. Warna titik setiap *retailer* diberikan berdasarkan hasil *clustering*. Untuk membuat visualisasi dengan plot 3D rgl, maka Script 5.13 yang dijalankan untuk 2 segmen dan 3 segmen.

```
#Grafik 3dimensi
output$thewidget1 <- renderRglwidget ({
  so <- pilihdata()
  open3d()
  plot3d(x=so$recency, y=so$frequency,z=so$monetary,
        col = warna[so$klasterke], type = 's', size = 1,
        xlab = "recency", ylab = "frequency",
        zlab = "monetary",
        colkey = list(length = 0.5, width = 0.5,
                      cex.clab = 0.75))
  scene1 <- scene3d()
  rgl.close()
  save <- options(rgl.inShiny = TRUE)
  on.exit(options(save))
  rglwidget(scene1)}})
```

Script 5.13 Visualisasi dengan grafik 3 dimensi pada 2 segmen dan 3 segmen

5.6.2.2. Visualisasi dengan Peta Lokasi

Proses pembuatan peta membutuhkan *package leaflet* dan data yang digunakan yaitu data hasil *clustering*, data longitude, dan latitude. Popup keterangan pada peta memerlukan data mengenai nama *retailer*, nilai *recency*, nilai *frequency* dan nilai *monetary* masing-masing *retailer*. Tampilan peta dunia akan lebih dibesarkan sehingga hanya terlihat pada kota Surabaya. Setiap titik akan menunjukkan letak *retailer* berdasarkan data longitude dan latitude. Warna titik setiap *retailer* diberikan berdasarkan hasil *clustering*. Keterangan setiap titik ditampilkan dengan *popup* dan muncul hanya jika titik diklik. Untuk membuat visualisasi dengan *leaflet*, maka Script 5.14 yang dijalankan pada halaman 3 segmen dan 2 segmen.

```
#Peta Lokasi
output$map<- renderLeaflet({
  dat <- pilihdata()
  leaflet() %>%
    addTiles() %>%
    setView(112.73439800,-7.28916600, zoom = 12) %>%
    addCircleMarkers(data = dat, ~lng, ~lat,
                    radius = 3,
                    color = warna[dat$klasterke],
                    popup = paste(
                      "<b>", dat$retailer,"</b>","<br>",
                      "Recency : ", dat$recency,"hari","<br>",
                      "Frequency : ", dat$frequency,"kali","<br>",
                      "Monetary : ", "Rp", dat$monetary )))})
```

Script 5.14 Visualisasi dengan peta lokasi pada 3 segmen dan 2 segmen

5.6.2.3. Visualisasi dengan Box Plot

Proses pembuatan box plot membutuhkan *package* *plotly* dan data yang digunakan yaitu data hasil *clustering* dan nilai *recency*, *frequency* *monetary*. Sumbu y menampilkan nilai RFM sedangkan sumbu x merupakan jenis segmen. Warna setiap box plot diberikan berdasarkan hasil *clustering*. Untuk membuat visualisasi dengan *box plot*, maka Script 5.15 yang dijalankan untuk menampilkan salah satu *box plot* yaitu untuk variabel nilai *recency* pada 3 segmen.

```
#Data box plot
y1 <- subset(ta.data,ta.data$klasterke==1)
y2 <- subset(ta.data,ta.data$klasterke==2)
y3 <- subset(ta.data,ta.data$klasterke==3)

#Box Plot
output$boxr <- renderPlotly ({
  plot_ly(type = 'box') %>%
    add_boxplot(y = y1$recency,
               name = "1",
               marker = list(color = "#D2691E"),
               line = list(color = "#D2691E")) %>%
    add_boxplot(y = y2$recency,
               name = "2",
               marker = list(color = "#1E90FF"),
               line = list(color = "#1E90FF")) %>%
    add_boxplot(y = y3$recency,
               name = "3",
               marker = list(color = "#DC143C"),
               line = list(color = "#DC143C"))
```

Script 5.15 Visualisasi dengan *box plot* 3 segmen

5.6.2.4. Visualisasi untuk Peringkat Retailer

Proses pembuatannya membutuhkan data code, nama *retailer*, dan nilai CLV setiap *retailer*. Nama *retailer* terbaik dan terburuk setiap segmen dicari menggunakan query dengan fungsi MIN dan MAX pada kolom nilai clv, lalu fungsi SELECT untuk menampilkan kolom nama *retailer* pada *valuebox*. Untuk membuat visualisasi peringkat *retailer*, maka Script 5.16 yang dijalankan untuk menampilkan salah satu *valuebox* yaitu *retailer* terbaik.

Pada tabel untuk menampilkan peringkat clv setiap *retailer*, kolom yang ditampilkan yaitu code, nama *retailer* dan nilai CLV. Banyak baris yang bisa ditampilkan yaitu 8 baris, namun

juga dapat menampilkan banyak baris sebanyak 5 atau 10 atau 15. Script 5.17 dijalankan untuk menampilkan tabel.

```
#Retailer Terbaik
output$baik<- rendervalueBox({
  re <- pilihdata()
  cari <- sqldf("select retailer From
                (select retailer,max(clv) From re)"
                )
  valueBox(
    value = tags$h5( "RETAILER TERBAIK"),
    subtitle = tags$h4(tags$strong (cari)),
    icon = icon("thumbs-o-up"),
    color = "purple"
  )})

#Retailer Terburuk
output$buruk<- rendervalueBox({
  re <- pilihdata()
  cari <- sqldf("Select retailer From
                (select retailer,min(clv) From re)")
  valueBox(
    value = tags$h5( "RETAILER TERBURUK"),
    subtitle = tags$h4(tags$strong (cari)),
    icon = icon("thumbs-o-down"),
    color = "purple"
  )})
```

Script 5.16 Visualisasi dengan *valuebox* untuk peringkat *retailer*

```
#CLV
output$clv <- renderDataTable({
  tabel <- pilihdata()
  kluaran <- data.frame(tabel$code,
                       tabel$retailer,
                       tabel$clv
  )
  colnames(kluaran) <- c("Code",
                        "Retailer",
                        "CLV")
  print(kluaran)
}, list(lengthMenu = c(5, 10, 15), pageLength = 8))
```

Script 5.17 Visualisasi dengan tabel untuk peringkat *retailer*

5.6.2.5. Visualisasi untuk Rentang RFM

Proses pembuatannya membutuhkan nilai *recency*, nilai *frequency* dan nilai *monetary* masing-masing *retailer*. Rentang ditampilkan dengan list yang berisi nilai minimal dan maksimal *recency*, *frequency*, *monetary* pada setiap segmennya. Untuk membuat visualisasi rentang RFM, maka Script 5.18 yang dijalankan.

```
#rentang Nilai RFM
output$range <- renderPrint({
  rentang <- pilihdata()
  list(Recency = paste(min(rentang$recency), "-",
                        max(rentang$recency), "hari"),
       Frequency = paste (min(rentang$frequency), "-",
                          max(rentang$frequency), "kali"),
       Monetary = paste ("Rp", min(rentang$monetary), "-",
                        "Rp", max(rentang$monetary))
  )
})
```

Script 5.18 Visualisasi untuk rentang RFM

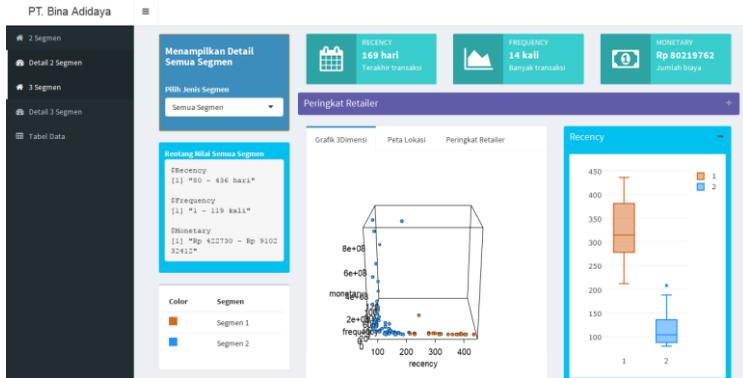
5.6.2.6. Visualisasi untuk Rata-rata RFM

Proses pembuatannya membutuhkan nilai *recency*, *frequency* dan *monetary* setiap *retailer*. Nilai rata-rata ditampilkan pada *infobox* yang berisi pembulatan dari perhitungan rata-rata *recency*, *frequency* dan *monetary* pada setiap segmennya. Fungsi MEAN dan ROUND digunakan untuk menghitung nilai rata-rata yang dibulatkan. Untuk membuat visualisasi rata-rata RFM, maka Script 5.19 dijalankan untuk menampilkan salah satu rata-rata yaitu *recency*.

```
#Rata-rata Recency
output$recency <- renderInfoBox({
  data <- pilihdata()
  infoBox(
    title = "Recency",
    value = paste(round(mean(data$recency)) , "hari"),
    subtitle = "Terakhir transaksi",
    icon = icon("calendar"),
    color = "teal",
    fill = TRUE))
#Rata-rata Frequency
output$frequency <- renderInfoBox({
  data <- pilihdata()
  infoBox(
    title = "Frequency",
    value = paste(round(mean(data$frequency)) , "kali"),
    subtitle = "Banyak transaksi",
    icon = icon("area-chart"),
    color = "teal",
    fill = TRUE))
#Rata-rata Monetary
output$monetary <- renderInfoBox({
  data <- pilihdata()
  infoBox(
    title = "Monetary",
    value = paste("Rp", round(mean(data$monetary))),
    subtitle = "Jumlah biaya",
    icon = icon("money"),
    color = "teal",
    fill = TRUE))
```

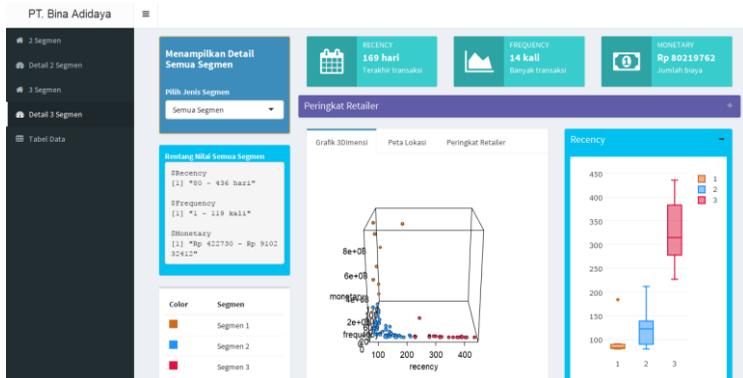
Script 5.19 Visualisasi untuk rata-rata RFM

Gambar 5.3 merupakan hasil tampilan visualisasi halaman detail 2 segmen berdasarkan pada desain antarmuka yang dibuat sebelumnya.



Gambar 5.3 Halaman detail 2 segmen

Gambar 5.4 merupakan hasil tampilan visualisasi halaman detail 3 segmen berdasarkan pada desain antarmuka yang dibuat sebelumnya.



Gambar 5.4 Halaman detail 3 segmen

5.7.3. Halaman Tabel data

Tabel data ditampilkan dalam visualisasi untuk menunjukkan data yang digunakan dalam proses segmentasi. Nama kolom yang ditampilkan yaitu data code, nama *retailer*, *recency*,

frequency monetary, nilai CLV dan hasil *clustering* masing-masing *retailer* ketika dibagi menjadi 2 segmen dan 3 segmen. Banyak baris yang bisa ditampilkan yaitu 10 baris, namun juga dapat menampilkan banyak baris sebanyak 5 atau 10 atau 15. Untuk membuat visualisasi tabel, maka Script 5.20 yang dijalankan.

```
#Data Tabel
output$mytable = renderDataTable({
  tabel <- alldata
  kluaran <- data.frame(tabel$code,
                       tabel$retailer,
                       tabel$recency,
                       tabel$frequency,
                       tabel$monetary,
                       tabel$clv,
                       tabel$ddua,
                       tabel$tiga
                      )

  colnames(kluaran) <- c("code",
                        "Retailer",
                        "Recency",
                        "Frequency",
                        "Monetary",
                        "CLV",
                        "2 Segmen",
                        "3 Segmen"
                       )

  print(kluaran)
})
list(lengthMenu = c(5, 10, 15), pageLength = 10))
```

Script 5.20 Visualisasi tabel data

Gambar 5.5 merupakan hasil tampilan visualisasi berdasarkan pada desain antarmuka yang dibuat sebelumnya.

Code	Retailer	Recency	Frequency	Monetary	CLV	2 Segmen	3 Segmen
GAI.0002	ANUGERAH	95	30	34128380	0.36903	2	2
GAI.0016	JAVA SANTOSA	87	18	80760900	0.30542	2	2
GAI.0018	GALLERY CAT / CV CAHAYA BINTANG SEJATI	83	76	285112802	0.67462	2	1
GAI.0020	KARYA SUBUR	391	1	10533050	0.02617	1	3
GAI.0021	MAJU	125	5	4565880	0.19915	2	2
GAI.0028	PUTRA JAYA	243	16	132215160	0.21027	1	3
GAI.0033	SAMUDRA JAYA	80	13	19769420	0.27390	2	2
GAI.0036	SATRIA	80	18	27677760	0.30468	2	2
GAI.0037	SEJAHTERA	90	19	30986820	0.30537	2	2
GAI.0042	SINAR JAYA II	128	1	844800	0.17307	2	2

Gambar 5.5 Halaman tabel data

BAB VI HASIL PEMBAHASAN

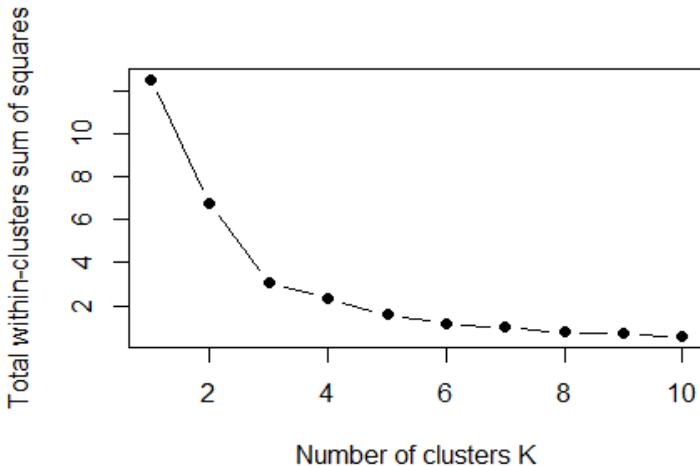
Pada bab ini akan menjelaskan hasil dan pembahasan dari proses *clustering* K-Means, uji performa *clustering*, pemberian bobot dan perhitungan CLV. Bab ini juga berisi analisis hasil proses *clustering* dan visualisasi hasil *clustering*.

6.1. Proses *Clustering*

Hasil dari proses implementasi dalam menentukan nilai k dengan metode Elbow dan melakukan *clustering* dengan metode K-Means akan dijelaskan lebih rinci pada setiap aktivitas.

6.1.1. Penentuan Nilai k dengan Metode Elbow

Plot hasil implementasi metode Elbow digambarkan seperti pada Gambar 6.1. Pada plot tersebut didapatkan titik siku yang terbentuk diantara titik dua dan empat, setelah titik 3 sudah tidak lagi terjadi penurunan yang signifikan, sehingga dapat disimpulkan bahwa jumlah *cluster* menurut metode Elbow yaitu sebanyak 3 *cluster*.



Gambar 6.1 Grafik hasil penerapan metode Elbow

6.1.2. Clustering dengan K-Means

Hasil dari proses *clustering* K-Means berupa informasi yang menunjukkan jumlah anggota setiap *cluster*, titik pusat atau centroid, dan nilai performa *cluster* berdasarkan *within cluster sum of squares*. *Clustering vector* menunjukkan jenis *cluster* pada setiap *retailer*, namun untuk lebih mudah dipahami hasil *clustering* ini perlu disimpan kedalam file dan hasilnya akan dibahas lebih rinci pada sub bab 6.6 mengenai analisis *cluster* yang terbentuk. Hasil dari proses penyimpanan data hasil *clustering* terdapat dalam Lampiran A.

6.1.2.1 Hasil K-Means 2 Cluster

Gambar 6.2 merupakan hasil dari 2 segmen jumlah anggota setiap *cluster* yang tidak sama banyak. *Cluster* yang pertama sebanyak 22 anggota dan *cluster* yang kedua sebesar 61 anggota. Pada 2 segmen ini terdapat dua titik centroid yang saling berjauhan.

```
k-means clustering with 2 clusters of sizes 22, 61
Cluster means:
  adidaya.rbalik adidaya.fnorm adidaya.mnorm
1      0.3296731      0.01155624      0.009282096
2      0.9007184      0.14656849      0.115991921

Clustering vector:
 [1] 2 2 2 1 2 1 2 2 2 2 2 1 1 2 2 2 2 2 2 2 2 1 2
 [26] 1 2 2 2 2 2 2 2 2 2 2 2 1 2 2 1 1 2 1 2 2 2 2 2
 [51] 2 1 1 2 1 1 1 1 2 2 2 2 1 2 2 2 1 2 2 2 2 2 2 2
 [76] 1 2 2 1 1 1 2 2

within cluster sum of squares by cluster:
 [1] 0.8005455 5.9835440
 (between_ss / total_ss = 45.9 %)
```

Gambar 6.2 Hasil *clustering* K-Means untuk 2 *cluster*

6.1.2.2 Hasil K-Means 3 Cluster

Gambar 6.3 merupakan hasil dari 3 segmen terdapat jumlah anggota setiap *cluster* yang tidak sama banyak karena terdapat *cluster* yang hanya memiliki 8 anggota dan terdapat *cluster* yang memiliki anggota terbanyak yaitu sebesar 54 anggota.

Pada 3 segmen ini terdapat titik titik centroid yang saling berjauhan terutama diantara titik centroid *cluster*1 dan titik centroid di *cluster* 3.

```

K-means clustering with 3 clusters of sizes 8, 54, 21

Cluster means:
  adidaya.rbalik  adidaya.fnorm  adidaya.mnorm
1      0.9540028      0.49258475   0.643374498
2      0.8877965      0.09274953   0.035781998
3      0.3154093      0.01170299   0.009547113

Clustering vector:
 [1] 2 2 1 3 2 3 2 2 2 2 3 3 1 2 1 2 1 2 1 1 2 2 3 2 3
    2 2 2 1 2 2 2 1 2 2 2 3 2 2 2 3 2 3 2 2 2 2 2 2
 [50] 2 2 3 3 2 3 3 3 3 2 2 2 2 3 2 2 2 2 3 2 2 2 2 2 2
     3 2 2 3 3 3 2 2

within cluster sum of squares by cluster:
 [1] 1.2552484 1.1152526 0.7065061
     (between_SS / total_SS =  75.5 %)

```

Gambar 6.3 Hasil *clustering* K-Means untuk 3 *cluster*

6.2. Verifikasi Hasil *Clustering*

Hasil verifikasi ini menunjukkan bahwa *retailer* yang menjadi satu segmen dari proses *clustering* K-Means merupakan *retailer* yang memiliki jarak terdekat ke centroid *cluster*. Titik centroid yang digunakan merupakan hasil dari proses *clustering* K-Means. Perhitungan jarak ke centroid untuk *retailer* secara keseluruhan terdapat pada Lampiran C.

6.2.1. Verifikasi 2 Segmen

Tabel 6.2 merupakan perhitungan jarak setiap *retailer* ke titik centroid telah terbukti bahwa lebih memilih jarak yang terpendek. Pembuktian pada *retailer* Anugerah yang merupakan anggota dari *cluster* 2, jika dihitung titik *retailer* ke dua centroid *cluster*, hasilnya jarak terpendek menggunakan perhitungan *Euclidean Distance* yaitu sebesar 0.139. *Retailer* Anugerah lebih sesuai berada di *cluster* 2 dibandingkan *cluster* lainnya.

Tabel 6.1 Potongan hasil perhitungan *Euclidean Distance* untuk 2 segmen

<i>Retailer</i>	K-Means <i>Cluster</i>	Jarak ke Centroid		Min	Hasil
		1	2		
Anugerah	2	0.671	0.139	0.139	2
Jaya Santosa	2	0.669	0.084	0.084	2
Gallery Cat / Cv. Cahaya Bintang Sejati	2	0.959	0.535	0.535	2
Karya Subur	1	0.204	0.795	0.204	1

6.2.2. Verifikasi 3 Segmen

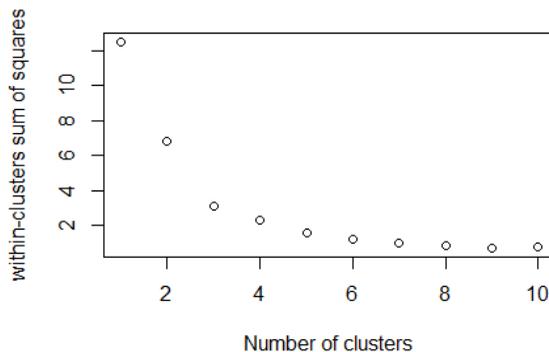
Tabel 6.2 merupakan perhitungan jarak setiap *retailer* ke titik centroid telah terbukti bahwa lebih memilih jarak yang terpendek. Pembuktian pada *retailer* Jaya Santosa yang merupakan anggota dari *cluster* 2, jika dihitung titik *retailer* ke tiga centroid *cluster*, hasilnya jarak terpendek menggunakan perhitungan *Euclidean Distance* yaitu sebesar 0.118. Jaya Santosa lebih sesuai berada di *cluster* 2 dibandingkan *cluster* lainnya.

Tabel 6.2 Potongan hasil perhitungan *Euclidean Distance* untuk 3 segmen

<i>Retailer</i>	K-Means <i>Cluster</i>	Jarak ke Centroid			Min	Hasil
		1	2	3		
Anugerah	2	0.655	0.168	0.684	0.168	2
Jaya Santosa	2	0.656	0.118	0.683	0.118	2
Gallery Cat / Cv. Cahaya Bintang Sejati	1	0.362	0.618	0.969	0.362	1
Karya Subur	3	1.152	0.767	0.189	0.189	3

6.3. Uji Performa *Clustering*

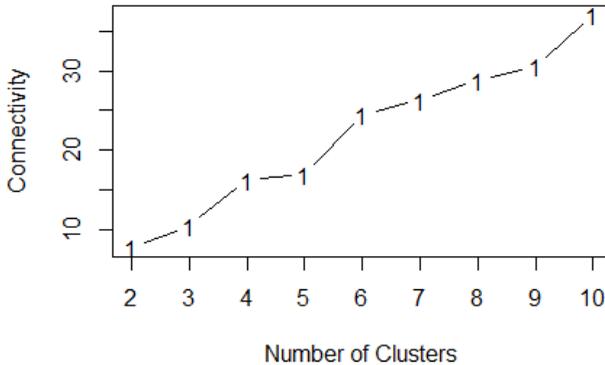
Pada pengujian ini akan mencari performa setiap jumlah *cluster* yang disesuaikan dengan rentang nilai pada metode Elbow. Pada metode SSE menampilkan grafik yang berisi nilai SSE disetiap percobaan jumlah *cluster* antara 2 hingga 10. Gambar 6.4 merupakan plot yang dihasilkan untuk jumlah segmen sebanyak 3 nilai SSEnya sebesar 3.0770, nilai tersebut bukan nilai SSE terendah namun merupakan nilai yang lebih rendah daripada jumlah segmen sebanyak 2.



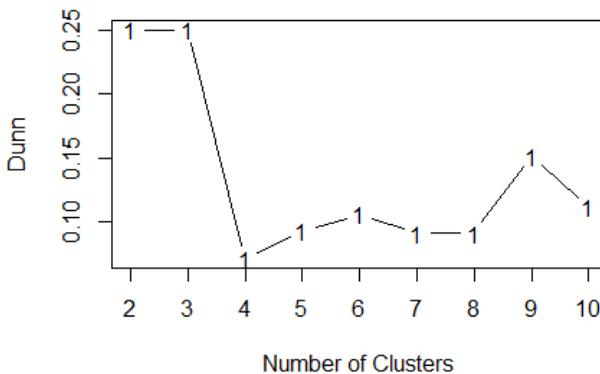
Gambar 6.4 Grafik hasil uji performa dengan metode SSE

Klaster	SSE
1	12.5354
2	6.7841
3	3.0770
4	2.4638
5	1.8585
6	1.5859
7	0.9919
8	0.9025
9	0.7990
10	0.7994

Pengujian hasil *clustering* juga dilakukan dengan metode *Connectivity* dan metode *Dunn Index*. Metode *Connectivity* dan *Dunn Index* menampilkan nilai setiap perhitungan berdasarkan jumlah *cluster* antara 2 hingga 10, sesuai dengan percobaan ketika menggunakan metode Elbow. Hasil *Connectivity* memilih jumlah *cluster* sebanyak 2 segmen karena memiliki nilai terendah diantara yang lain yaitu sebesar 7.8933 dan hasil *Dunn Index* juga memilih jumlah *cluster* sebanyak 2 segmen karena memiliki nilai tertinggi diantara yang lain yaitu sebesar 0,2503. Plot yang dihasilkan untuk metode *Connectivity* pada Gambar 6.5 dan *Dunn Index* ditunjukkan pada Gambar 6.6.



Gambar 6.5 Grafik hasil uji performa dengan metode *Connectivity*



Gambar 6.6 Grafik hasil uji performa dengan metode *Dunn Index*

Nilai setiap *cluster* menggunakan metode *Connectivity* dan *Dunn Index* untuk jumlah *cluster* dari 2 hingga 10, direkap seperti pada Tabel 6.3 untuk mengetahui perbedaan nilai yang dimiliki masing-masing metode. Pengujian performa *clustering* dengan menggunakan tiga metode terdapat perbedaan namun membuat dapat lebih mengetahui dari 10 percobaan terdapat 2 pilihan jumlah *cluster* yaitu antara jumlah segmen sebanyak 2 dan 3. Pilihan jumlah *cluster* berdasarkan metode Elbow sebanyak 3 tidak sesuai dengan metode *Dunn Index* dan *Connectivity*, sehingga melakukan proses *clustering* kembali dengan menggunakan jumlah nilai *k* sebanyak 2 dan menganalisis perbedaan karakteristik kelompok yang dihasilkan antara jumlah segmen sebanyak 2 dan 3 serta memilih jumlah segmen yang terbaik diantara keduanya.

Tabel 6.3 Rekapitulasi hasil uji performa *clustering* dari semua metode

Klaster	Connectivity	Dunn
1		
2	7.8933	0.2503
3	10.5234	0.2503
4	16.2845	0.0719
5	16.9623	0.0936
6	24.5381	0.1061
7	26.3337	0.0916
8	28.8337	0.0916
9	30.5337	0.1515
10	36.9397	0.1118

6.4. Pemberian Bobot RFM

Proses perhitungan AHP dilakukan untuk proses awal dalam menganalisis peringkat *cluster* dan *retailer* terbaik. Hasil bobot yang didapatkan seperti pada Tabel 6.4. Berdasarkan proses AHP dapat diketahui bahwa *frequency* merupakan kriteria terpenting yang digunakan untuk menilai *retailer* yang terbaik.

Setelah *frequency*, penilaian akan mengutamakan kriteria *recency* dibandingkan dengan kriteria *monetary*. Kriteria *monetary* menjadi penilaian paling akhir dalam penentuan *retailer* terbaik. Penentuan besarnya bobot ini masih memiliki nilai CI sebesar 0.034 dan nilai tidak sama dengan 0, sehingga bobot yang didapatkan masih berdasarkan pada pengisian kuisisioner yang belum konsisten. Namun tingkat ketidakkonsistensinya masih dapat diterima karena memiliki nilai CI/RI lebih kecil dari 0.1 yaitu sebesar 0.058.

Tabel 6.4 Bobot setiap kriteria

Kriteria	Bobot
Recency	0.20
Frequency	0.71
Monetary	0.09

6.5. Perhitungan CLV

Perhitungan CLV menghasilkan peringkat *cluster* yang memiliki nilai tertinggi hingga terendah. Nilai indeks CLV dengan nilai terbesar menunjukkan bahwa *retailer* dalam *cluster* tersebut merupakan *retailer* yang memiliki tingkat loyalitas yang paling tinggi dibandingkan *cluster* lainnya.

6.5.1. Perhitungan CLV 2 Segmen

Tabel 6.5 merupakan peringkat setiap *cluster* untuk jumlah segmen sebanyak 2. *Cluster* 1 merupakan peringkat kedua karena index CLVnya memiliki nilai *recency*, *frequency* dan *monetary* yang lebih rendah dibandingkan *cluster* 2.

Tabel 6.5 CLV untuk jumlah segmen sebanyak 2

Cluster	Indeks CLV	Peringkat CLV
1	0.07488	2
2	0.29349	1

6.5.2. Perhitungan CLV 3 Segmen

Tabel 6.6 merupakan peringkat setiap *cluster* untuk jumlah segmen sebanyak 3. *Cluster* 1 merupakan peringkat pertama karena index CLVnya memiliki nilai *recency*, *frequency* dan *monetary* yang lebih lebih besar dibandingkan *cluster* 2 dan 3, sedangkan *cluster* 3 merupakan peringkat terendah.

Tabel 6.6 CLV untuk jumlah segmen sebanyak 3

<i>Cluster</i>	Indeks CLV	Peringkat CLV
1	0.59201	1
2	0.24627	2
3	0.07215	3

6.6. Analisis Cluster yang Terbentuk

Sebanyak 83 *retailer* yang menjadi pelanggan PT. Bina Adidaya terbagi ke dalam dua dan tiga segmen. Setiap segmen memiliki karakter yang berbeda-beda, secara lebih rinci karakteristik kelompok pelanggan yang terbentuk akan dijelaskan pada setiap aktivitas.

6.6.1. Analisis Karakteristik 2 Segmen

Tabel 6.9 menunjukkan bahwa *retailer* yang termasuk dalam *cluster* 1 ini merupakan *retailer* yang memiliki peringkat lebih rendah dari *cluster* 2. Rata-rata nilai *recency* sebesar 319 hari menunjukkan bahwa *retailer* melakukan transaksi terakhir pada waktu yang sudah cukup lama, bukan pada bulan-bulan terakhir tahun 2016. Jika melihat dari rentang nilai *cluster*, *retailer* melakukan transaksi sekitar pada bulan Januari hingga bulan Agustus saja di tahun 2016. Nilai rata-rata *frequency* yang sebesar 2 sampai 3 kali menunjukkan bahwa *retailer* hanya melakukan sedikit sekali transaksi di tahun 2016. Nilai rata-rata *monetary* yang sebesar 8.867.670 rupiah menunjukkan bahwa *retailer* hanya menghabiskan sedikit uangnya untuk membeli produk PT. Bina Adidaya di tahun 2016.

Tabel 6.7 Karakteristik *cluster* 1 ketika 2 segmen

Peringkat <i>Cluster</i>	Kedua
<i>Cluster</i>	1
Jumlah Anggota	22
Karakteristik	<i>Recency</i> : 212 – 436 hari <i>Frequency</i> : 1 – 16 kali <i>Monetary</i> : 422.730 – 132.215.160 rupiah

Tabel 6.10 menunjukkan bahwa *retailer* yang termasuk dalam *cluster* 2 ini merupakan *retailer* yang memiliki peringkat lebih tinggi dari *cluster* 1. Nilai rata-rata *recency* yang sebesar 115 menunjukkan bahwa *retailer* melakukan transaksi terakhir pada pada bulan-bulan terakhir tahun 2016. Jika melihat dari rentang nilai *cluster*, *retailer* masih aktif melakukan transaksi ketika bulan Agustus hingga Desember. Nilai rata-rata *frequency* yang sebesar 18 kali menunjukkan bahwa *retailer* melakukan transaksi satu sampai dua kali setiap bulan di tahun 2016. Nilai rata-rata *monetary* yang sebesar 105.953.303 rupiah menunjukkan bahwa *retailer* yang termasuk cukup besar karena dapat menghabiskan banyak uangnya untuk membeli produk PT. Bina Adidaya.

Tabel 6.8 Karakteristik *cluster* 2 ketika 2 segmen

Peringkat <i>Cluster</i>	Pertama
<i>Cluster</i>	2
Jumlah Anggota	61
Karakteristik	<i>Recency</i> : 80 – 208 hari <i>Frequency</i> : 1 – 119 kali <i>Monetary</i> : 657.360 – 910.232.412 rupiah

6.6.2. Analisis Karakteristik 3 Segmen

Tabel 6.9 menunjukkan bahwa *retailer* yang termasuk dalam *cluster* 1 ini merupakan *retailer* yang memiliki peringkat tertinggi. Dilihat dari rata-rata nilai *recency*, *frequency* dan *monetary* yang dimiliki *cluster* 1 ini lebih baik dibandingkan

cluster lainnya. Nilai rata-rata *recency* yang sebesar 96 hari menunjukkan bahwa *retailer* melakukan transaksi terakhir pada pada bulan-bulan terakhir tahun 2016. Jika melihat dari rentang nilai *cluster*, *retailer* masih aktif melakukan transaksi ketika bulan September hingga bulan Desember. Nilai rata-rata *frequency* yang sebesar 59 kali menunjukkan bahwa *retailer* melakukan transaksi setiap bulannya bisa lebih dari sekali di tahun 2016. Nilai rata-rata *monetary* yang sebesar 585.771.078 rupiah menunjukkan bahwa *retailer* yang termasuk besar karena mampu menghabiskan banyak uangnya untuk membeli produk PT. Bina Adidaya di tahun 2016.

Tabel 6.9 Karakteristik *cluster* 1 ketika 3 segmen

Peringkat <i>Cluster</i>	Pertama
<i>Cluster</i>	1
Jumlah Anggota	8
Karakteristik	<i>Recency</i> : 80 – 184 hari <i>Frequency</i> : 23 – 119 kali <i>Monetary</i> : 166.625.800 – 910.232.412 rupiah

Tabel 6.10 menunjukkan bahwa *retailer* yang termasuk dalam *cluster* 2 ini merupakan *retailer* yang memiliki peringkat cukup tinggi. Dilihat dari rata-rata nilai *recency*, *frequency* dan *monetary* yang dimiliki *cluster* 2 ini lebih baik dibandingkan *cluster* 3 namun tidak lebih baik dari *cluster* 1. Nilai rata-rata *recency* yang sebesar 120 hari menunjukkan bahwa *retailer* melakukan transaksi terakhir pada pada bulan-bulan terakhir tahun 2016. Jika melihat dari rentang nilai *cluster*, *retailer* masih aktif melakukan transaksi ketika bulan Agustus hingga bulan Desember. Nilai rata-rata *frequency* yang sebesar 12 kali menunjukkan bahwa *retailer* melakukan transaksi hampir pada setiap bulannya di tahun 2016. Nilai rata-rata *monetary* yang sebesar 32.977.538 rupiah menunjukkan bahwa *retailer* yang termasuk menengah karena dapat menghabiskan cukup banyak uangnya untuk membeli produk PT. Bina Adidaya di tahun 2016.

Tabel 6.10 Karakteristik *cluster 2* ketika 3 segmen

Peringkat <i>Cluster</i>	Kedua
<i>Cluster</i>	2
Jumlah Anggota	54
Karakteristik	<i>Recency</i> : 80 – 212 hari <i>Frequency</i> : 1 – 43 kali <i>Monetary</i> : 657.360 - 237.539.940 rupiah

Tabel 6.11 menunjukkan bahwa *retailer* yang termasuk dalam *cluster 3* ini merupakan *retailer* yang memiliki peringkat terendah. Dilihat dari rata-rata nilai *recency*, *frequency* dan *monetary* yang dimiliki *cluster 3* tidak begitu baik dibandingkan dengan dua *cluster* lainnya. Nilai rata-rata *recency* yang sebesar 324 hari menunjukkan bahwa *retailer* melakukan transaksi terakhir pada waktu yang sudah cukup lama, bukan pada bulan-bulan terakhir tahun 2016. Jika melihat dari rentang nilai *cluster*, *retailer* melakukan transaksi sekitar pada bulan Januari hingga bulan April saja di tahun 2016. Nilai rata-rata *frequency* yang sebesar 2 sampai 3 kali menunjukkan bahwa *retailer* hanya melakukan sedikit sekali transaksi di tahun 2016. Nilai rata-rata *monetary* yang sebesar 9.108.786 rupiah menunjukkan bahwa *retailer* hanya menghabiskan sedikit uangnya untuk membeli produk PT. Bina Adidaya di tahun 2016.

Tabel 6.11 Karakteristik *cluster 3* ketika 3 segmen

Peringkat <i>Cluster</i>	Ketiga
<i>Cluster</i>	3
Jumlah Anggota	21
Karakteristik	<i>Recency</i> : 227 – 436 hari <i>Frequency</i> : 1 – 16 kali <i>Monetary</i> : 422.730 - 132.215.160 rupiah

6.6.3. Analisis Persamaan dan Perbedaan 2 Segmen dan 3 Segmen

Pada anggota *cluster* 1 ketika menjadi 2 segmen dengan *cluster* 3 ketika menjadi 2 segmen, hanya terdapat satu perbedaan anggota *retailer* yaitu Anugrah. Jika *cluster* 1 dan 2 ketika menjadi 3 segmen digabung, maka anggota kelompok hanya terdapat satu perbedaan anggota yaitu *retailer* Anugrah. *Retailer* Anugrah ini memiliki nilai *recency* sebesar 212 hari, *frequency* sebanyak 2 kali dan *monetary* sebesar 3.804.240 rupiah. Nilai *recency* yang dimiliki Anugrah menjadi batas minimal rentang nilai di *cluster* 2 ketika menjadi 3 segmen dan menjadi batas maksimal rentang nilai di *cluster* 1 ketika menjadi 2 segmen. Kesimpulan yang didapatkan dari analisis jumlah anggota dan nilai RFM *retailer*, menunjukkan bahwa pembagian menjadi 2 segmen cenderung melihat pada satu variabel yaitu *recency* dan terbukti dengan tidak ada *overlapping* pada pembagian kelompok jika dilihat dari nilai *recency*. Penambahan kelompok yang dihasilkan untuk pembagian menjadi 3 segmen yaitu dengan memisahkan menjadi dua kelompok pada anggota *cluster* 2 ketika menjadi 2 segmen dan memindahkan *retailer* yang memiliki nilai *recency* tertinggi pada *cluster* 1 ketika menjadi 2 segmen ke dalam anggota *cluster* 2 ketika menjadi 3 segmen.

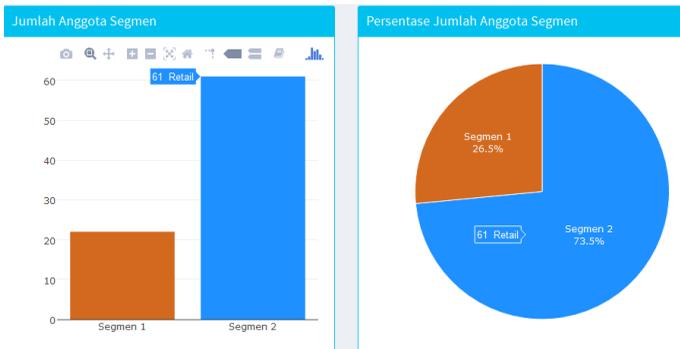
6.7. Analisis berdasarkan Visualisasi

Hasil segmentasi ditampilkan dalam bentuk grafik untuk membantu pengambilan keputusan dari perusahaan dalam menganalisis *retailer* dari PT. Bina Adidaya. Hasil visualisasi setiap pembagian segmen 2 dan segmen 3 akan dijelaskan lebih rinci pada setiap aktivitas.

6.7.1. Analisis Visualisasi 2 Segmen

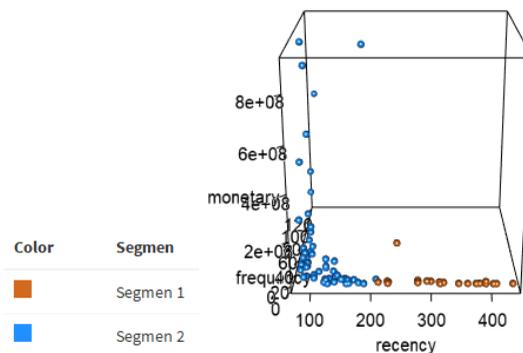
Gambar 6.7 menunjukkan grafik *bar chart* dan *pie chart* jumlah anggota setiap segmennya. Segmen 1 digambarkan dengan warna kuning memiliki anggota lebih sedikit dibandingkan segmen 2 yaitu sebesar 26.5%. Segmen 2 digambarkan dengan

warna biru memiliki anggota lebih banyak dibandingkan segmen 1 yaitu sebesar 73.5%.



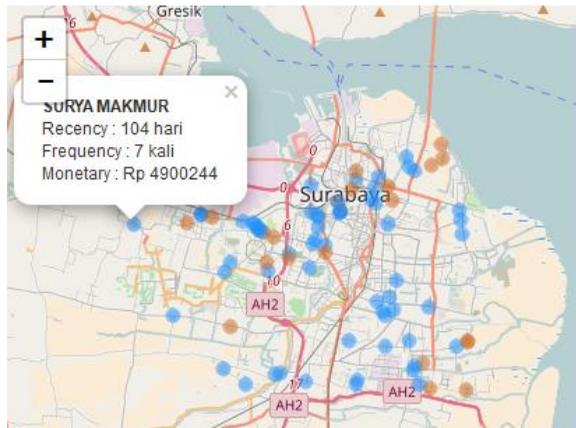
Gambar 6.7 Grafik bar chart dan pie chart untuk 2 segmen

Gambar 6.8 menunjukkan persebaran 83 *retailer* yang akan dikelompokkan menjadi 2 segmen. Informasi yang didapatkan bahwa *retailer* yang termasuk dalam segmen 1 memiliki nilai *recency* yang menjauhi 0, nilai *frequency* yang mendekati 0, dan nilai *monetary* yang mendekati 0. Titik kelompok pada segmen 1 terdapat satu *retailer* yang terpisah dari kelompok titik lainnya karena memiliki nilai *frequency* dan *monetary* yang cukup jauh dari 0. *Retailer* yang termasuk dalam segmen 2 memiliki nilai *recency* yang mendekati 0, nilai *frequency* yang menyebar, dan nilai *monetary* yang juga menyebar. Segmen 2 kelompok titik lebih menyebar dari segmen 1.



Gambar 6.8 Grafik 3 dimensi untuk 2 segmen

Gambar 6.9 menunjukkan persebaran lokasi 83 *retailer* di Surabaya. Didapatkan informasi yaitu *retailer* yang menjadi satu segmen tidak berkumpul pada satu kawasan tertentu, namun menyebar di berbagai kawasan di Surabaya.



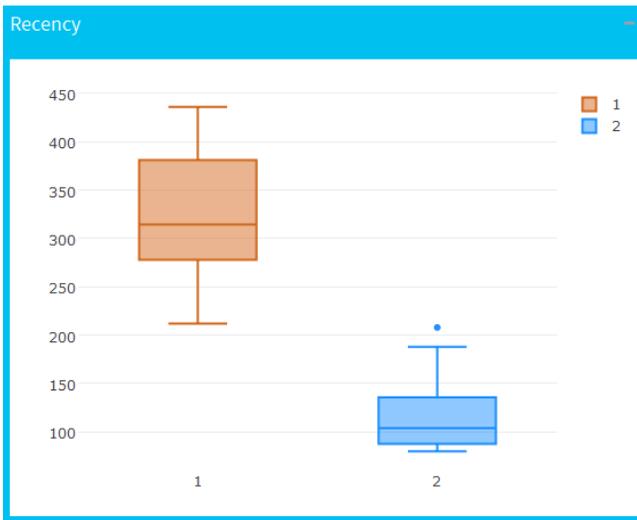
Gambar 6.9 Peta lokasi untuk 2 segmen

Gambar 6.10 merupakan contoh inputan segmen 1 yang menunjukkan bahwa Putra Jaya merupakan *retailer* terbaik dan UD. Rukun Jaya merupakan *retailer* terburuk di segmen 1.



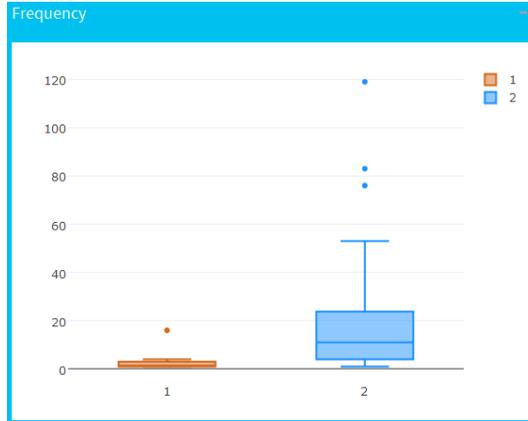
Gambar 6.10 Peringkat *retailer* segmen 1 untuk 2 segmen

Gambar 6.11 merupakan boxplot *recency* kedua segmen. Segmen 1 memiliki bentuk boxplot dan whisker yang lebih panjang dan terletak diatas box lainnya, sehingga rentang nilai yang terbentuk lebih panjang daripada segmen 1. Segmen 2 memiliki boxplot dan whisker yang lebih kecil, rentang nilai *recency* yang terbentuk menjadi melebar karena terdapat outlier jauh diatas whisker. Rentang nilai segmen 1 dan segmen 3 sangat baik karena terpisah satu dengan yang lainnya. *Recency* di segmen 2 jauh lebih baik dibandingkan segmen 1 karena sebagian *retailer* memiliki nilai terendah yang tidak beragam.



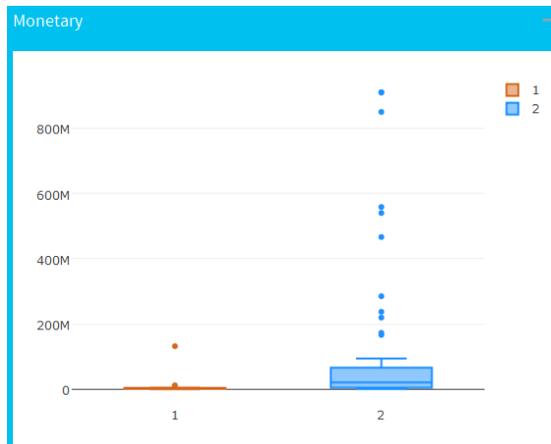
Gambar 6.11 Box plot *recency* untuk 2 segmen

Gambar 6.12 merupakan boxplot *frequency* kedua segmen. Segmen 1 memiliki bentuk boxplot yang lebih pendek, sehingga rentang nilai yang terbentuk juga pendek. Segmen 2 memiliki boxplot dan whisker yang panjang dan memiliki outlier, sehingga rentang nilai yang terbentuk menjadi sangat panjang. Segmen 1 terletak diantara kuartil atas segmen 2 dan nilai terendah segmen 2, sehingga rentang nilai segmen 1 merupakan bagian dari rentang nilai segmen 2. *Frequency* di segmen 2 jauh lebih baik dibandingkan segmen 1 karena sebagian *retailer* memiliki nilai tertinggi namun beragam.



Gambar 6.12 Box plot *frequency* untuk 2 segmen

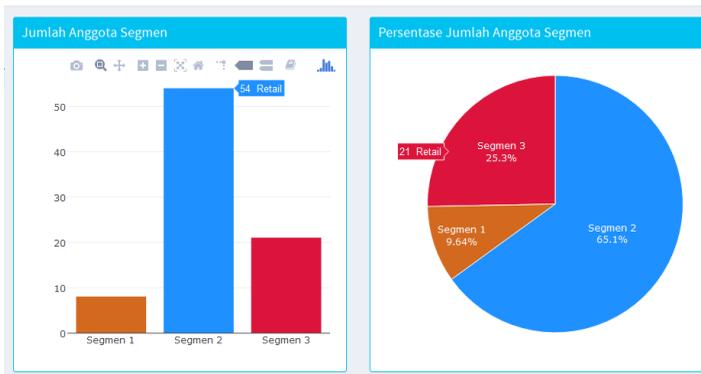
Gambar 6.13 merupakan boxplot *monetary* kedua segmen. Segmen 1 memiliki bentuk boxplot yang lebih pendek, sehingga rentang nilai yang terbentuk juga pendek. Segmen 2 memiliki boxplot dan whisker yang panjang dan memiliki outlier yang sangat banyak, sehingga rentang yang terbentuk menjadi sangat panjang. Rentang nilai segmen 1 merupakan bagian dari rentang nilai segmen 2. *Monetary* segmen 2 jauh lebih baik dibandingkan segmen 1 karena sebagian *retailer* memiliki nilai sangat tinggi dan beragam.



Gambar 6.13 Box plot *monetary* untuk 2 segmen

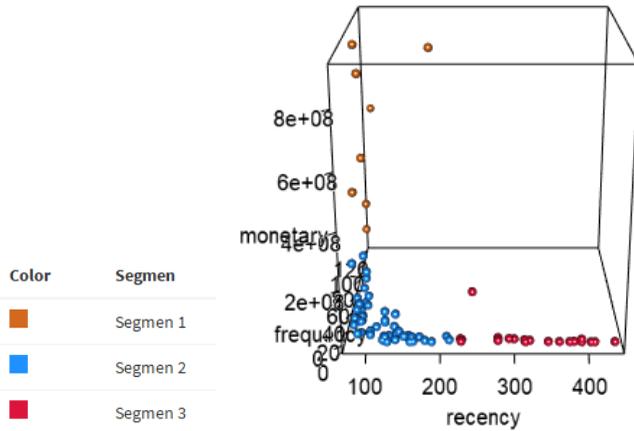
6.7.2. Analisis Visualisasi 3 Segmen

Gambar 6.14 menunjukkan grafik *bar chart* dan *pie chart* jumlah anggota setiap segmennya. Segmen 1 digambarkan dengan warna kuning memiliki anggota paling sedikit dibandingkan segmen lainnya yaitu sebesar 9.6%. Segmen 2 digambarkan dengan warna biru memiliki anggota paling banyak dibandingkan segmen lainnya yaitu sebesar 65.1%. Segmen 3 digambarkan dengan warna merah memiliki anggota sebesar 25.3%.



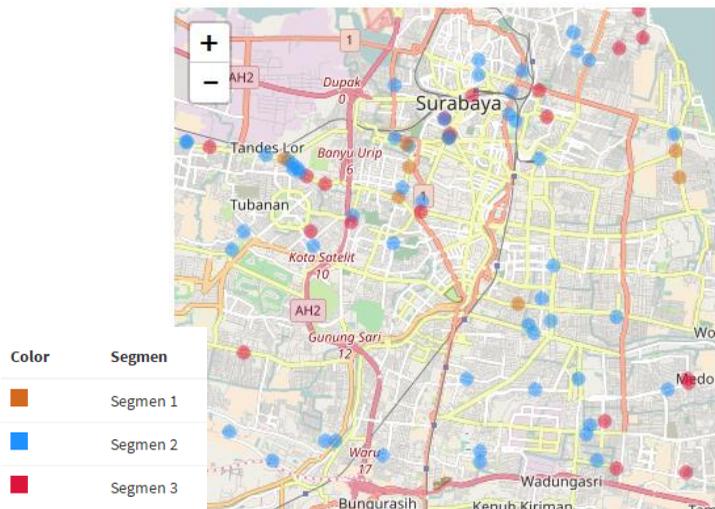
Gambar 6.14 Grafik *bar chart* dan *pie chart* untuk 3 segmen

Gambar 6.15 menunjukkan persebaran 83 *retailer* yang akan dikelompokkan menjadi 3 segmen. Informasi yang didapatkan bahwa *retailer* yang termasuk dalam segmen 1 memiliki nilai *recency* yang mendekati 0, nilai *frequency* yang menjauhi 0, dan nilai *monetary* yang menjauhi 0. Titik kelompok pada segmen 1 tidak terlalu rapat dibandingkan segmen lainnya. *Retailer* yang termasuk dalam segmen 2 memiliki nilai *recency*, *frequency* dan *monetary* yang mendekati 0. Segmen 2 kelompok titik lebih rapat dan tidak terdapat *retailer* yang terpisah jauh. *Retailer* tersebut memiliki nilai *recency* yang menjauhi 0, nilai *frequency* yang mendekati 0, dan nilai *monetary* yang mendekati 0. Segmen 3 terdapat satu *retailer* yang terpisah dari kelompok titik lainnya karena memiliki nilai *frequency* dan *monetary* yang cukup jauh dari 0.



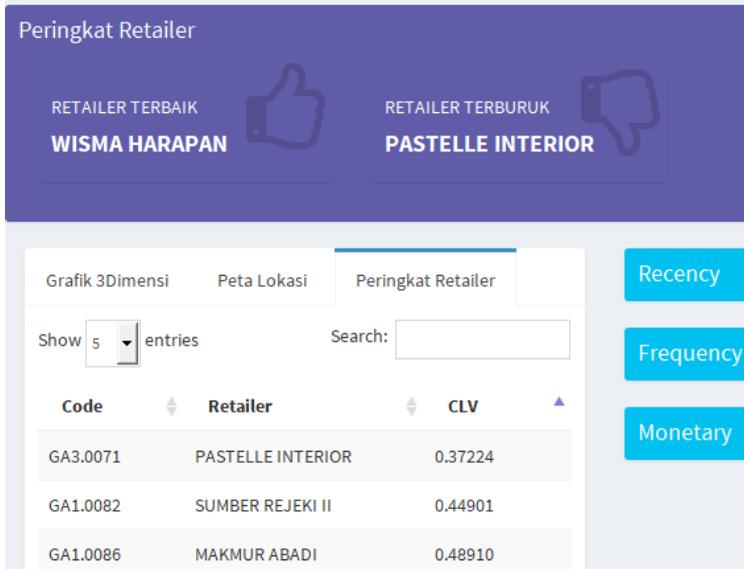
Gambar 6.15 Grafik 3 dimensi untuk 3 segmen

Gambar 6.16 menunjukkan persebaran lokasi 83 *retailer* di Surabaya. Didapatkan informasi yaitu *retailer* yang menjadi satu segmen tidak berkumpul pada satu kawasan tertentu, namun menyebar di berbagai kawasan di Surabaya.



Gambar 6.16 Peta lokasi untuk 3 segmen

Gambar 6.17 merupakan contoh inputan segmen 1 yang menunjukkan bahwa Wisma Harapan merupakan *retailer* terbaik dan Pastelle Interior merupakan *retailer* terburuk di segmen 1.



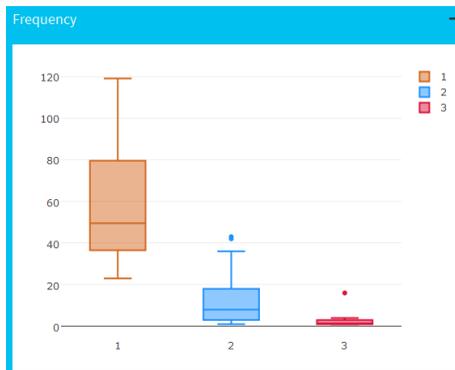
Gambar 6.17 Peringkat *retailer* dengan inputan semua segmen

Gambar 6.18 merupakan boxplot *recency* ketiga segmen. Segmen 1 memiliki bentuk boxplot dan whisker yang lebih kecil dan terletak dibawah box lainnya, rentang nilai *recency* yang terbentuk menjadi melebar karena terdapat outlier jauh diatas whisker. Rentang nilai yang terbentuk pada segmen 1 merupakan bagian nilai terendah dari segmen 2, sehingga menyebabkan rentang nilai beririsan. Segmen 3 memiliki boxplot dan whisker yang panjang, sehingga rentang yang terbentuk paling lebar dari segmen lainnya. Rentang nilai segmen 3 lebih baik karena terpisah dengan boxplot lainnya. *Recency* di segmen 1 jauh lebih baik dibandingkan lainnya karena sebagian *retailer* memiliki nilai terendah yang tidak beragam.



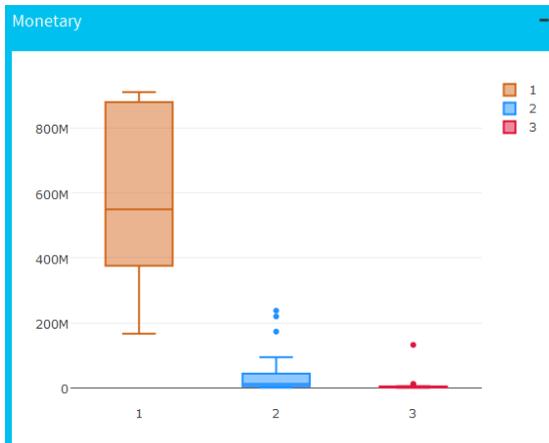
Gambar 6.18 Box plot *recency* untuk 3 segmen

Gambar 6.19 merupakan boxplot *frequency* ketiga segmen. Segmen 1 memiliki bentuk boxplot yang lebih panjang, sehingga rentang nilai yang terbentuk juga panjang. Segmen 1 terletak diatas boxplot segmen 1 tetapi terdapat nilai di bawah median yang sama dengan nilai tertinggi dan outlier segmen 2, sehingga segmen 1 memiliki rentang yang beririsan dengan segmen 2. Segmen 3 memiliki boxplot dan whisker yang pendek namun memiliki outlier, sehingga rentang yang terbentuk sama dengan nilai terendah hingga kuartil teratas segmen 2 tetapi lebih rentang segmen 3 lebih sempit dari lainnya. *Frequency* di segmen 1 jauh lebih baik dibandingkan lainnya karena sebagian *retailer* memiliki nilai tertinggi namun beragam.



Gambar 6.19 Box plot *frequency* untuk 3 segmen

Gambar 6.20 merupakan boxplot *monetary* ketiga segmen. Segmen 1 memiliki bentuk boxplot yang lebih panjang dan terletak diatas boxplot lainnya, sehingga rentang nilai *monetary* yang terbentuk lebih lebar. Segmen 2 memiliki nilai outlier yang terletak pada nilai terendah di segmen 1, sehingga rentang nilai segmen 1 dan 2 saling berisihan. Segmen 3 memiliki boxplot dan whisker yang pendek namun terdapat outlier sehingga rentangnya menjadi lebih lebar tetapi paling sempit dari yang lainnya. Rentang nilai segmen 3 hampir sama dengan segmen 2 hanya berbeda sedikit di rentang nilai teratas karena outlier segmen 2 lebih banyak dan bernilai tinggi. *Monetary* segmen 1 jauh lebih baik dibandingkan lainnya karena sebagian *retailer* memiliki nilai tertinggi namun beragam.



Gambar 6.20 Box plot *monetary* untuk 3 segmen

6.8. Kesimpulan Analisis

Hasil uji performa *cluster* menggunakan metode *Dunn Index* dan *Connectivity* yang memilih jumlah *cluster* menjadi 2, sehingga dapat memisahkan *retailer* yang melakukan transaksi pada bulan Januari hingga minggu ketiga Agustus dengan *retailer* yang melakukan transaksi pada minggu terakhir Agustus hingga bulan Desember. Hasil 2 segmen tersebut dapat membantu perusahaan memisahkan *retailer* yang tergolong masih aktif dan sudah mulai tidak aktif.

Karakteristik yang terbentuk dari pembagian menjadi 2 segmen yaitu *retailer* yang termasuk dalam *cluster* 1 ini memberikan kontribusi yang sangat sedikit untuk perusahaan, sehingga *cluster* 1 merupakan *retailer* yang tidak perlu dipertahankan oleh perusahaan. *Cluster* 2 dinilai mampu berkontribusi untuk perusahaan di masa depan, sehingga *retailer* yang termasuk dalam *cluster* 2 perlu dipertahankan oleh perusahaan.

Hasil penentuan jumlah *cluster* berdasarkan metode Elbow yang menentukan jumlah *cluster* menjadi 3, meskipun masih terdapat nilai variabel yang *overlapping* tetapi penentuan jumlah *cluster* ini dapat memisahkan kelompok *retailer* yang tergolong masih aktif dan cenderung memiliki nilai *recency*, *frequency* dan *monetary* yang lebih unggul dari kelompok lain.

Karakteristik yang terbentuk dari pembagian menjadi 3 segmen yaitu *cluster* 1 dinilai mampu memberikan kontribusi keuntungan yang sangat besar di masa depan, sehingga *retailer* yang termasuk dalam *cluster* 1 sangat perlu dipertahankan oleh perusahaan. Kontribusi yang diberikan *retailer* pada *cluster* 2 tidak terlalu besar dibandingkan dengan *cluster* 1, akan tetapi perusahaan perlu membuat strategi yang dapat menjadikan *cluster* 2 untuk cenderung naik dalam memberikan kontribusi keuntungan. *Retailer* yang termasuk dalam *cluster* 3 memberikan kontribusi yang sangat sedikit untuk perusahaan, sehingga *cluster* 3 merupakan *retailer* yang paling terakhir menjadi prioritas pelayanan atau bahkan menjadi *retailer* yang tidak perlu diperhatikan lebih oleh perusahaan.

Perhitungan CLV dapat membantu menunjukkan segmen *retailer* yang terbaik dan terburuk serta membantu memberikan peringkat kepada *retailer* secara keseluruhan berdasarkan pertimbangan kriteria yang lebih mementingkan *frequency* dari transaksi yang dilakukan *retailer*.

Berdasarkan analisis terhadap karakteristik RFM setiap segmen yang terbentuk dari jumlah segmen sebanyak 2 dan jumlah segmen sebanyak 3, rentang nilai yang terbentuk lebih baik menggunakan jumlah segmen sebanyak 3 karena pada jumlah segmen sebanyak 2 yaitu di variabel *frequency* dan *monetary*

tidak dapat menunjukkan karakteristik dengan tepat dan rentang nilainya lebih *overlapping*. Jika merujuk nilai SSE yang lebih minimum, maka segmen sebanyak 3 lebih baik daripada segmen sebanyak 2. Jika melihat dari keterbatasan sumber daya perusahaan untuk mengelolah *retailer* akan lebih mudah mempertahankan *retailer* yang memiliki karakteristik nilai RFM terbaik dan termasuk dalam kelompok terbaik.

BAB VII KESIMPULAN DAN SARAN

Bab ini memaparkan kesimpulan yang diperoleh dari rangkaian pengerjaan tugas akhir yang telah dilakukan serta menyertakan saran untuk pengembangan tugas akhir dengan topik serupa.

7.1. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa metode K-Means dapat digunakan sebagai pilihan dalam menyelesaikan masalah segmentasi pelanggan untuk mengetahui kelompok *retailer*. Hasil penerapan metode K-Means telah diuji performa menggunakan *Dunn Index* dan *Connectivity*. Hasil uji tersebut belum sesuai dengan hasil pencarian nilai k yang menggunakan metode Elbow sehingga dilakukan *clustering* dua kali. Model RFM yang dihasilkan dapat mendeskripsikan karakteristik setiap segmen dibuktikan dengan dapat diketahuinya perilaku setiap *retailer* dalam melakukan transaksi kepada PT. Bina Adidaya. Hubungan antara model RFM dengan metode K-Means dapat membentuk segmentasi pelanggan yang sesuai dengan keadaan *retail* karena nilai RFM *retailer* yang berbeda-beda menjadi data yang diolah dalam metode K-Means, sehingga dapat dikelompokkan menjadi satu segmen dengan melihat kedekatan titik nilai RFM setiap *retailer* dan dapat mengetahui *retailer* mana yang perlu dipertahankan.

Visualisasi berbasis *web-based* ditampilkan lebih interaktif yaitu menggabungkan beberapa grafik dan isi grafik dapat disesuaikan dengan pilihan jenis segmen yang diinginkan. Visualisasi grafik *pie chart* dan *bar chart* menunjukkan jumlah pembagian anggota setiap segmennya, persebaran *retailer* dalam dilihat pada grafik 3 dimensi dan peta lokasi, karakter setiap segmen yang ditunjukkan melalui range nilai RFM, divisualisasikan melalui box plot, dan terakhir dapat melihat peringkat *retailer* terbaik dan terburuk disetiap segmen.

7.2. Saran

Saran yang dapat dipertimbangkan untuk pengembangan tugas akhir ini yaitu:

- a. Penelitian ini dapat dikembangkan dengan menambahkan seluruh *retailer* yang berada di kota manapun dan menjadi pelanggan PT. Bina Adidaya sebagai studi kasus.
- b. Data transaksi penjualan produk PT. Bina Adidaya yang digunakan sebaiknya ditambahkan rentang waktu lebih dari 1 tahun.
- c. Kategori produk pada jenis pemasaran proyek sebaiknya ditambahkan untuk merepresentasikan seluruh jenis pelanggan yang dimiliki PT. Bina Adidaya.
- d. Penambahan variabel juga dapat dilakukan untuk pengembangan penelitian dan mempertajam analisis. Contoh variabel yang dapat digunakan yaitu demografi mengenai jenis pelanggan, apakah termasuk dalam pelanggan pada pemasaran retail atau proyek
- e. Penelitian ini juga dapat dikembangkan dengan menggabungkan metode lainnya untuk membantu mencari nilai k yang lebih optimal ketika proses *clustering* K-Means.
- f. Penelitian dapat dikembangkan dengan perbandingan hasil segmentasi dengan menggunakan metode *clustering* lainnya ataupun dengan mengkombinasikan metode K-Means dengan metode *clustering* lainnya.

DAFTAR PUSTAKA

- [1] V. A. Muliana, “Indonesia Jadi Pasar Empuk Produk Cat Luar Negeri,” *Liputan6*, 20 Oktober 2016. [Online]. Available: <http://bisnis.liputan6.com/read/2631095/indonesia-jadi-pasar-empuk-produk-cat-luar-negeri>. [Diakses 10 Januari 2017].
- [2] U. Kulsum dan S. Cicilia, “Pasar bisnis cat tetap menjanjikan,” *Kontan.co.id*, 13 Oktober 2016. [Online]. Available: <http://industri.kontan.co.id/news/pasar-bisnis-cat-tetap-menjanjikan>. [Diakses 10 Januari 2017].
- [3] Ke Lu dan T. Furukawa, “A Framework for Segmenting Customers based on Probability Density of Transaction Data,” *International Conference on Advanced Applied Informatics*, pp. 273-278, 2012.
- [4] D. Suryadi, “Strategi Avian Mewarnai Pasar Cat,” 19 September 2016. [Online]. Available: <http://swa.co.id/swa/trends/strategi-avian-mewarnai-pasar-cat>. [Diakses 11 Januari 2017].
- [5] B. Shim, K. Choi dan Y. Suh, “CRM strategies for a small-sized online shopping mall based on association rules and sequential patterns,” *Expert Systems with Applications*, p. 7736–7742, 2012.
- [6] M. Mohammadian dan I. Makhani, “RFM-Based customer segmentation as an elaborative analytical tool for enriching the creation of sales and trade marketing strategies,” *International Academic Journal of Accounting and Financial Management*, vol. 3, pp. 21-35., 2016.
- [7] Aviliani, U. Sumarwan, I. Sugema dan A. Saefuddin, “Segmentasi Nasabah Tabungan Mikro Berdasarkan Recency, Frequency, dan Monetary : Kasus Bank BRI,” pp. 95-109.

- [8] K. Tsipstis dan A. Chorianopoulos, *Data Mining Techniques in CRM: Inside Customer Segmentation*, United Kingdom: John Wiley and Sons, Ltd, 2009.
- [9] Jing Wu dan Zheng Lin , “Research on Customer Segmentation Model by Clustering,” pp. 316-318, 2005.
- [10] M. J. A. Berry dan G. S. Linoff, *Mastering Data Mining The Art and Science of Customer Relationship Management*, Canada: John Wiley & Sons, Inc, 2000.
- [11] J. Oscar Ong, “Implementasi Algoritma K-Means Clustering untuk Menentukan Strategi Marketing President University,” *Jurnal Ilmiah Teknik Industri*, vol. 12, pp. 10-20, 2013.
- [12] P. N. Tan, M. Steinbach dan V. Kumar, *Introduction to Data Mining*, Boston: Addison-Wesley Longman Publishing Co, 2005.
- [13] T. M. Kodinariya dan D. P. R. Makwana, “Review on determining number of Cluster in K-Means Clustering,” *International Journal of Advance Research in Computer Science and Management Studies*, vol. 1, pp. 90-95, 2013.
- [14] J. . T. Wei, Y.-Z. Yang, S.-Y. Lin dan H.-H. Wu, “Applying Data Mining and RFM Model to Analyze Customers’ Values of A Veterinary Hospital,” *International Symposium on Computer, Consumer and Control*, pp. 481-484, 2016.
- [15] A. Dursun dan M. Caber, “Using Data Mining Techniques for Profiling Profitable Hotel Customers: an Application of RFM Analysis,” *Tourism Management Perspectives*, vol. 18, p. 153–160, 2016.
- [16] M. Khajvand dan M. . J. Tarokh, “Estimating Customer Future Value of Different Customer Segments Based on Adapted RFM Model in Retail Banking Context,” *Procedia Computer Science*, vol. 3, p. 1327–1332, 2011.
- [17] S. M. Seyed Hosseini, A. Maleki dan M. R. Gholamian, “Cluster Analysis Using Data Mining Approach to Develop CRM Methodology to Assess the Customer

- Loyalty,” *Expert Systems with Applications*, vol. 37, p. 5259–5264, 2010.
- [18] C.-H. Cheng dan Y.-S. Chen, “Classifying the segmentation of customer value via RFM model and RS theory,” *Expert Systems with Applications*, vol. 36, p. 4176–4184, 2009.
- [19] J. Winser dan L. Stanley, *Process Management: Creating Value Along the Supply Chain* 1st edition.
- [20] W. Wagner dan M. Zubey, *Customer Relationship Management a People Process and Technology Approach*, United State of America: Thomson Course Technology, 2007.
- [21] R. Kalakota dan M. Robinson, *E - Business Roadmap for Success Second Edition*, Amerika Serikat: Addison Wesley Longman, 2001.
- [22] S. M. H. Jansen, “Customer Segmentation and Customer Profiling,” 2007.
- [23] J. Han, M. Kamber dan J. Pei, *Data Mining Concepts and Techniques Third Edition*, San Fransisco: Morgan Kauffman, 2006.
- [24] C.-Y. Chiu, Y.-F. Chen, I.-T. Kuo dan H. C. Ku, “An intelligent market segmentation system using k-means and particle swarm optimization,” *Expert Systems with Applications*, p. 4558–4565, 2009.
- [25] C. Fraley dan A. E. Fraley, “Technical Report No. 329,” dalam *How Many Clusters? Which Clustering Method? Answers*, Department of Statisticstics University of Washington, 1998.
- [26] O. Maimon dan L. Rokach, *Data Mining and Knowledge Discovery Handbo- ok*, Secaucus: Springer-Verlag New York, Inc, 2005.
- [27] J. D. Banfield dan A. E. Raftery, “Model-based Gaussian and non-Gaussian clustering, *Biometrics*,” 1993, pp. 49:803-821.

- [28] N. Wakhidah , “Clustering Menggunakan K-Means Algorithm (K-Means Algorithm Clustering)”.
- [29] P. Bholowalia dan A. Kumar, “EBK-Means: A Clustering Techniques based on Elbow Method and K-Means in WSN,” *International Journal of Computer Applications*, vol. 105, pp. 17-24, 2014.
- [30] S. G. Krishna Patro dan K. K. Sahu, “Normalization: A Preprocessing Stage”.
- [31] H. Junaedi, H. Budianto, I. Maryati dan Y. Melani, “Data Transformation Pada Data Mining,” *Prosiding Konferensi Nasional “Inovasi dalam Desain dan Teknologi”*, pp. 93-99, 2011.
- [32] D. P. Dewi, P. A. Djunaidy, M.Sc., Ph.D, M.T dan R. Pradina,S.T, “Evaluasi dan Rekomendasi Peningkatan Nilai Hidup Pelanggan Berdasarkan Analisis RFM Perilaku Pembelian Pelanggan : Studi Kasus PT. XYZ,” *Jurnal Teknik POMITS*, vol. 2, 2013.
- [33] E. Martiana, “Data Preprocessing,” Soft Computation Research Group, EEPIS-ITS, Surabaya, 2013.
- [34] S. Saitta, B. Raphael dan I. F. C. Smith, “A Bounded Index for Cluster Validity,” *Machine Learning and Data Mining in Pattern Recognition*, pp. 174-187, 2007.
- [35] J. K. Fluegemann, M. D. Davies dan N. D. Aguirre, “Determining the Optimal Number of Clusters with the Clustergram,” *NASA USRP – Internship Final Report*, 2011.
- [36] G. Brock, V. Pihur, S. Datta dan S. Datta, “clValid, an R package for cluster validation,” *Journal of Statistical Softwar*, 2011.
- [37] Y.-S. Chen, C.-H. Cheng, . C.-J. Lai, C.-Y. Hsu dan . H.-J. Syu, “Identifying patients in target customer segments using a two-stage clustering-classification approach: A hospital-based assessment,” *Computers in Biology and Medicine*, p. 213–221, 2012.

- [38] A. M. Hughes, dalam *Strategic database marketing*, Chicago, Probus Publishing Company.
- [39] R. J. Baron dan R. J. Galka, *CRM The Foundation of Contemporary Marketing Strategy*, UK: Routledge, 2013.
- [40] J.-T. Wei, S.-Y. Lin dan H.-H. Wu, "A review of the application of RFM model," *African Journal of Business Management*, vol. 4, pp. 4199-4206, 2010.
- [41] B. W. Taylor, *Introduction to Management Science*, United States of America: Pearson Education, Inc, 2013.
- [42] S. M. Rezaeinia, A. Keramati dan A. Albadvi , "An Integrated AHP-RFM Method to Banking Customer Segmentation," *Electronic Customer Relationship Management*, vol. 6, pp. 153-168, 2012.
- [43] S. Gupta dan D. Lehmann, "Customer as Assets," *Journal of Interactive Marketing*, vol. 17, pp. 9-24, 2003.

Halaman ini sengaja dikosongkan

BIODATA PENULIS



Penulis lahir di Surabaya pada tanggal 25 bulan April tahun 1995 ini merupakan anak pertama dari dua bersaudara. Penulis telah menempuh pendidikan formal di SD Negeri Jemur Wonosari I Surabaya, SMP Negeri 12 Surabaya, dan SMA Negeri 1 Surabaya. Pasca kelulusan penulis pada jenjang SMA, penulis melanjutkan pendidikan ke jenjang perguruan tinggi di

Departemen Sistem Informasi, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember (ITS) dan terdaftar sebagai mahasiswa Sistem Informasi angkatan 2013. Selama menjadi mahasiswa, penulis aktif dalam kegiatan kemahasiswaan dan mengikuti kepanitiaan di tingkat departemen dan institusi. Penulis aktif dalam berorganisasi dengan menjadi staff ahli di Media Informasi, Himpunan Mahasiswa Sistem Informasi. Pada semester delapan perkuliahan, penulis mulai mengerjakan Tugas Akhir di Laboratorium Rekayasa Data dan Intelegensi Bisnis, di bawah bimbingan Bapak Edwin Riksakomara, S.Kom., M.T., penulis mengambil topik mengenai data mining. Penulis dapat dihubungi melalui email anissa.veronika.ava@gmail.com. Semoga penulisan Tugas Akhir ini mampu memberikan kontribusi positif bagi semua pihak terkait.

Halaman ini sengaja dikosongkan

Lampiran A
DATA TUGAS AKHIR

Tabel A.1 Data tugas akhir hasil pra proses yang digabungkan dengan hasil *clustering*

code	retailer	R	F	M	rnorm	rbalik	fnorm	mnorm	clv	2 segmen	3 segmen
GA1.0002	Anugerah	95	30	34128380	0.04	0.96	0.25	0.04	0.369	2	2
GA1.0016	Jaya Santosa	87	18	80760900	0.02	0.98	0.14	0.09	0.305	2	2
GA1.0018	Gallery Cat / CV. Cahaya Bintang Sejati	83	76	285132802	0.01	0.99	0.64	0.31	0.675	2	1
GA1.0020	Karya Subur	391	1	10533050	0.87	0.13	0.00	0.01	0.026	1	3
GA1.0021	Maju	125	5	4565880	0.13	0.87	0.03	0.00	0.199	2	2
GA1.0028	Putra Jaya	243	16	132215160	0.46	0.54	0.13	0.14	0.210	1	3

code	retailer	R	F	M	rnorm	rbalik	fnorm	mnorm	clv	2 segmen	3 segmen
GA1.0033	Samudra Jaya	80	13	19769420	0.00	1.00	0.10	0.02	0.274	2	2
GA1.0036	Satria	80	18	27677760	0.00	1.00	0.14	0.03	0.305	2	2
GA1.0037	Sejahtera	90	19	30980620	0.03	0.97	0.15	0.03	0.305	2	2
GA1.0042	Sinar Jaya II	128	1	844800	0.13	0.87	0.00	0.00	0.173	2	2
GA1.0044	Sinar Langgeng	151	1	21120000	0.20	0.80	0.00	0.02	0.162	2	2
GA1.0051	Subur II	361	1	1236840	0.79	0.21	0.00	0.00	0.042	1	3
GA1.0057	Surya	314	2	4359960	0.66	0.34	0.01	0.00	0.075	1	3
GA1.0065	Wisma Harapan	82	119	558604750	0.01	0.99	1.00	0.61	0.958	2	1
GA1.0069	Kharisma Jaya	163	1	1958000	0.23	0.77	0.00	0.00	0.154	2	2
GA1.0074	Sumber Rejeki	80	53	909026660	0.00	1.00	0.44	1.00	0.593	2	1

code	retailer	R	F	M	rnorm	rbalik	fnorm	mnorm	clv	2 segmen	3 segmen
GA1. 0081	CV. Sejahtera	124	10	79599300	0.12	0.88	0.08	0.09	0.236	2	2
GA1. 0082	Sumber Rejeki II	91	31	849738050	0.03	0.97	0.25	0.93	0.449	2	1
GA1. 0083	CV.Seng Joyo Mandiri	124	9	43845340	0.12	0.88	0.07	0.05	0.227	2	2
GA1. 0086	Makmur Abadi	89	42	540188864	0.03	0.97	0.35	0.59	0.489	2	1
GA1. 0096	J O Y O	184	46	910232412	0.29	0.71	0.38	1.00	0.492	2	1
GA3. 0007	Budhi Jaya	156	6	5784240	0.21	0.79	0.04	0.01	0.188	2	2
GA3. 0008	Dwi Tunggal	159	2	3481500	0.22	0.78	0.01	0.00	0.162	2	2
GA3. 0011	Jaya Abadi	278	2	1506450	0.56	0.44	0.01	0.00	0.095	1	3
GA3. 0014	Karunia Jaya	91	9	5834400	0.03	0.97	0.07	0.01	0.242	2	2
GA3. 0026	Naga Jaya	229	3	1117600	0.42	0.58	0.02	0.00	0.128	1	3

code	retailer	R	F	M	rnorm	rbalik	fnorm	mnorm	clv	2 segmen	3 segmen
GA3.0027	Naga Emas	171	5	4066260	0.26	0.74	0.03	0.00	0.173	2	2
GA3.0037	Putra Makmur	144	4	27704160	0.18	0.82	0.03	0.03	0.184	2	2
GA3.0038	Putra Mas	125	4	5286710	0.13	0.87	0.03	0.01	0.193	2	2
GA3.0044	Sinar Abadi	80	83	166625800	0.00	1.00	0.69	0.18	0.708	2	1
GA3.0045	Sinar Agung	88	22	44808390	0.02	0.98	0.18	0.05	0.326	2	2
GA3.0059	Warna	122	3	2897400	0.12	0.88	0.02	0.00	0.189	2	2
GA3.0068	Kota Jaya	124	2	1134430	0.12	0.88	0.01	0.00	0.181	2	2
GA3.0071	Pastelle Interior	82	23	466619284	0.01	0.99	0.19	0.51	0.372	2	1
GA3.0080	Dua Jaya Tunggal	87	42	60752602	0.02	0.98	0.35	0.07	0.448	2	2
GA3.0084	Surya Makmur	104	7	4900244	0.07	0.93	0.05	0.00	0.223	2	2

code	retailer	R	F	M	rnorm	rbalik	fnorm	mnorm	clv	2 segmen	3 segmen
GA3.0085	Agung Makmur	90	43	94237000	0.03	0.97	0.36	0.10	0.455	2	2
GA3.0097	G M (UD. Hm. Masri S.)	321	2	2009040	0.68	0.32	0.01	0.00	0.071	1	3
GA3.0142	Saputra Jaya	90	22	23319934	0.03	0.97	0.18	0.03	0.323	2	2
GA3.0160	Aneka Logam	208	6	2244000	0.36	0.64	0.04	0.00	0.158	2	2
GA3.0162	Anugrah	212	2	3804240	0.37	0.63	0.01	0.00	0.132	1	2
GA3.0165	Baru	278	4	5135581	0.56	0.44	0.03	0.01	0.107	1	3
GA3.0166	Gangsar	103	11	5189800	0.06	0.94	0.08	0.01	0.248	2	2
GA3.0167	Pacific Raya Motor	300	3	4037880	0.62	0.38	0.02	0.00	0.089	1	3
GA3.0169	Sinar Indah	139	2	4844400	0.17	0.83	0.01	0.00	0.173	2	2

code	retailer	R	F	M	rnorm	rbalik	fnorm	mnorm	clv	2 segmen	3 segmen
GA3.0170	Sumber Utama Karya	90	18	18712980	0.03	0.97	0.14	0.02	0.298	2	2
GA3.0173	Wahyu	90	14	29197960	0.03	0.97	0.11	0.03	0.275	2	2
GA3.0175	Sunrise	161	3	10747440	0.23	0.77	0.02	0.01	0.167	2	2
GA3.0177	Warna Jaya	125	4	5588880	0.13	0.87	0.03	0.01	0.193	2	2
GA3.0204	Subur	123	5	6256800	0.12	0.88	0.03	0.01	0.200	2	2
GA3.0205	Ultimate Shine	124	9	50272200	0.12	0.88	0.07	0.05	0.228	2	2
GA3.0211	Proline	404	1	471900	0.91	0.09	0.00	0.00	0.018	1	3
GA3.0212	Bandar Jaya	409	1	451440	0.92	0.08	0.00	0.00	0.015	1	3
GA3.0213	Habza	135	15	7199280	0.15	0.85	0.12	0.01	0.254	2	2
GA3.0214	Guna Jaya	375	1	855360	0.83	0.17	0.00	0.00	0.034	1	3

code	retailer	R	F	M	rnorm	rbalik	fnorm	mnorm	clv	2 segmen	3 segmen
GA3.0215	Pratama	278	3	2413620	0.56	0.44	0.02	0.00	0.101	1	3
GA3.0216	Mukti Jaya	227	3	4524080	0.41	0.59	0.02	0.00	0.130	1	3
GA3.0217	Makmur Jaya	346	1	1483240	0.75	0.25	0.00	0.00	0.051	1	3
GA3.0218	Mahkota Auto Paint	188	1	686400	0.30	0.70	0.00	0.00	0.139	2	2
GA3.0219	Leo	125	3	10926630	0.13	0.87	0.02	0.01	0.188	2	2
GA4.0002	Aa Elektro	90	36	25217830	0.03	0.97	0.30	0.03	0.407	2	2
GA4.0011	Bangun Rejeki	130	5	9633265	0.14	0.86	0.03	0.01	0.197	2	2
GA4.0023	Istana	391	1	422730	0.87	0.13	0.00	0.00	0.025	1	3
GA4.0054	Sinar Jaya	87	6	11406780	0.02	0.98	0.04	0.01	0.227	2	2
GA4.0067	Lidah Utomo	87	17	23995400	0.02	0.98	0.14	0.03	0.294	2	2

code	retailer	R	F	M	rnorm	rbalik	fnorm	mnorm	clv	2 segmen	3 segmen
GA4.0084	Tri Stars	179	2	1721940	0.28	0.72	0.01	0.00	0.151	2	2
GA4.0088	Subur Jaya III	315	1	558800	0.66	0.34	0.00	0.00	0.068	1	3
GA4.0093	Artomoro	91	30	219692000	0.03	0.97	0.25	0.24	0.388	2	2
GA4.0094	Valao	138	11	66818070	0.16	0.84	0.08	0.07	0.233	2	2
GA4.0097	Cahaya Makmur	139	7	7576690	0.17	0.83	0.05	0.01	0.204	2	2
GA4.0109	Hartono Jaya	96	32	63442610	0.04	0.96	0.26	0.07	0.383	2	2
GA4.0111	Yudi	111	13	12588840	0.09	0.91	0.10	0.01	0.256	2	2
GA4.0118	Prado	80	15	237539940	0.00	1.00	0.12	0.26	0.305	2	2
GA4.0119	Evan Jaya Makmur	80	31	59417270	0.00	1.00	0.25	0.06	0.386	2	2
GA4.0121	Adi Jaya	96	26	173120290	0.04	0.96	0.21	0.19	0.357	2	2

code	retailer	R	F	M	rnorm	rbalik	fnorm	mnorm	clv	2 segmen	3 segmen
GA4. 0139	A.A	228	1	2059200	0.42	0.58	0.00	0.00	0.117	1	3
GA4. 0141	Sumber Rejeki III	87	6	13064040	0.02	0.98	0.04	0.01	0.227	2	2
GA4. 0142	Teng Joyo	84	16	66225280	0.01	0.99	0.13	0.07	0.294	2	2
GA4. 0143	Inti Varian	381	1	2085380	0.85	0.15	0.00	0.00	0.031	1	3
GA4. 0146	UD. Rukun Jaya	436	1	1438800	1.00	0.00	0.00	0.00	0.000	1	3
GA4. 0148	CV. Beruang Sukses Makmur	293	1	12368400	0.60	0.40	0.00	0.01	0.081	1	3
GA4. 0149	Varian	122	2	3540790	0.12	0.88	0.01	0.00	0.183	2	2
GA4. 0150	Insan Jaya	159	1	657360	0.22	0.78	0.00	0.00	0.156	2	2

Halaman ini sengaja dikosongkan

Lampiran B ARIBUT DATA MENTAH

Tabel B.1 Atribut data mentah

Atribut	Keterangan
No	Nomor setiap baris dalam laporan
Product Code	Kode produk yang dipasarkan
Product Name	Nama produk yang dipasarkan
Packing atau Pack size	Ukuran berat pengemasan produk yang dipasarkan
SJ No	Nomor surat jalan
SJ Update	Tanggal surat jalan saat dibuat
Customer atau Debtor name	Nama <i>retailer</i>
Qty	Jumlah produk yang dibeli <i>retailer</i>
DO No	Nomor oder
Quantity LTR	Ukuran berat pengemasan produk yang dipasarkan dalam satuan liter
Quantity KG	Ukuran berat pengemasan produk yang dipasarkan dalam satuan kilogram
Price Per Unit Rp	Harga satuan produk yang dipasarkan dalam mata uang rupiah
Total Rp	Total harga perproduk yang dibeli <i>retailer</i> dalam mata uang rupiah
Code	Kode <i>retailer</i>
Address 1	Alamat tempat <i>retailer</i>
Address 2	Nomor telepon <i>retailer</i>
City	Kota tempat <i>retailer</i>

B-2

Halaman ini sengaja dikosongkan

Lampiran C
HASIL PERHITUNGAN EUCLIDEAN DISTANCE

Tabel C.1 hasil perhitungan *Euclidean Distance* untuk 3 segmen

<i>Retailer</i>	K-Means Cluster	Jarak ke Centroid			Min	Hasil
		1	2	3		
Anugerah	2	0.655	0.168	0.684	0.168	2
Jaya Santosa	2	0.656	0.118	0.683	0.118	2
Gallery Cat / CV. Cahaya Bintang Sejati	1	0.362	0.618	0.969	0.362	1
Karya Subur	3	1.152	0.767	0.189	0.189	3
Maju	2	0.791	0.068	0.559	0.068	2
Putra Jaya	3	0.743	0.364	0.288	0.288	3
Samudra Jaya	2	0.736	0.113	0.691	0.113	2
Satria	2	0.707	0.124	0.698	0.124	2
Sejahtera	2	0.698	0.103	0.672	0.103	2
Sinar Jaya II	2	0.815	0.102	0.550	0.102	2
Sinar Langgeng	2	0.807	0.128	0.485	0.128	2
Subur II	3	1.099	0.684	0.106	0.106	3
Surya	3	1.008	0.552	0.028	0.028	3
Wisma Harapan	1	0.510	1.081	1.343	0.510	1
Kharisma Jaya	2	0.830	0.156	0.452	0.156	2
Sumber Rejeki	1	0.362	1.030	1.277	0.362	1

C-2

<i>Retailer</i>	K-Means Cluster	Jarak ke Centroid			Min	Hasil
		1	2	3		
CV. Sejahtera	2	0.699	0.055	0.570	0.055	2
Sumber Rejeki II	1	0.376	0.916	1.158	0.376	1
CV.Seng Joyo Mandiri	2	0.736	0.030	0.565	0.030	2
Makmur Abadi	1	0.155	0.619	0.942	0.155	1
J O Y O	1	0.447	1.022	1.128	0.447	1
Budhi Jaya	2	0.798	0.117	0.472	0.117	2
Dwi Tunggal	2	0.822	0.142	0.463	0.142	2
Jaya Abadi	3	0.952	0.453	0.129	0.129	3
Karunia Jaya	2	0.766	0.090	0.656	0.090	2
Naga Jaya	3	0.882	0.318	0.266	0.266	3
Naga Emas	2	0.814	0.158	0.430	0.158	2
Putra Makmur	2	0.783	0.096	0.505	0.096	2
Putra Mas	2	0.795	0.075	0.558	0.075	2
Sinar Abadi	1	0.505	0.630	0.983	0.505	1
Sinar Agung	2	0.673	0.124	0.684	0.124	2
Warna	2	0.801	0.083	0.567	0.083	2
Kota Jaya	2	0.808	0.092	0.561	0.092	2
Pastelle Interior	1	0.335	0.497	0.863	0.335	1

<i>Retailer</i>	K-Means Cluster	Jarak ke Centroid			Min	Hasil
		1	2	3		
Dua Jaya Tunggal	2	0.596	0.273	0.747	0.273	2
Surya Makmur	2	0.777	0.069	0.618	0.069	2
Agung Makmur	2	0.558	0.284	0.747	0.284	2
G M (UD. Hm. Masri S.)	3	1.022	0.572	0.011	0.011	3
Saputra Jaya	2	0.694	0.120	0.677	0.120	2
Aneka Logam	2	0.844	0.255	0.327	0.255	2
Anugrah	2	0.865	0.274	0.314	0.274	2
Baru	3	0.941	0.450	0.129	0.129	3
Gangsar	2	0.758	0.057	0.624	0.057	2
Pacific Raya Motor	3	0.981	0.512	0.067	0.067	3
Sinar Indah	2	0.810	0.105	0.519	0.105	2
Sumber Utama Karya	2	0.714	0.100	0.670	0.100	2
Wahyu	2	0.722	0.086	0.664	0.086	2
Sunrise	2	0.812	0.140	0.457	0.140	2
Warna Jaya	2	0.795	0.075	0.558	0.075	2
Subur	2	0.788	0.066	0.564	0.066	2
Ultimate Shine	2	0.730	0.033	0.566	0.033	2
Proline	3	1.185	0.804	0.226	0.226	3

C-4

<i>Retailer</i>	K-Means Cluster	Jarak ke Centroid			Min	Hasil
		1	2	3		
Bandar Jaya	3	1.195	0.818	0.240	0.240	3
Habza	2	0.746	0.057	0.541	0.057	2
Guna Jaya	3	1.126	0.723	0.145	0.145	3
Pratama	3	0.947	0.452	0.129	0.129	3
Mukti Jaya	3	0.877	0.312	0.272	0.272	3
Makmur Jaya	3	1.071	0.643	0.064	0.064	3
Mahkota Auto Paint	2	0.850	0.215	0.382	0.215	2
Leo	2	0.795	0.081	0.558	0.081	2
Aa Elektro	2	0.647	0.221	0.716	0.221	2
Bangun Rejeki	2	0.788	0.070	0.545	0.070	2
Istana	3	1.158	0.768	0.190	0.190	3
Sinar Jaya	2	0.776	0.108	0.666	0.108	2
Lidah Utomo	2	0.714	0.102	0.677	0.102	2
Tri Stars	2	0.837	0.189	0.407	0.189	2
Subur Jaya III	3	1.017	0.557	0.029	0.029	3
Artomoro	2	0.472	0.269	0.732	0.269	2
Valao	2	0.711	0.063	0.531	0.063	2
Cahaya Makmur	2	0.783	0.073	0.520	0.073	2

<i>Retailer</i>	K-Means Cluster	Jarak ke Centroid			Min	Hasil
		1	2	3		
Hartono Jaya	2	0.618	0.186	0.690	0.186	2
Yudi	2	0.743	0.035	0.604	0.035	2
Prado	2	0.537	0.253	0.737	0.253	2
Evan Jaya Makmur	2	0.627	0.199	0.728	0.199	2
Adi Jaya	2	0.533	0.206	0.694	0.206	2
A.A	3	0.889	0.319	0.269	0.269	3
Sumber Rejeki III	2	0.774	0.108	0.666	0.108	2
Teng Joyo	2	0.679	0.113	0.686	0.113	2
Inti Varian	3	1.137	0.740	0.162	0.162	3
UD. Rukun Jaya	3	1.251	0.893	0.316	0.316	3
CV. Beruang Sukses Makmur	3	0.972	0.495	0.087	0.087	3
Varian	2	0.806	0.090	0.567	0.090	2
Insan Jaya	2	0.829	0.148	0.463	0.148	2

C-6

Tabel C.2 hasil perhitungan *Euclidean Distance* untuk 2 segmen

<i>Retailer</i>	K-Means Cluster	Jarak ke Centroid		Min	Hasil
		1	2		
Anugerah	2	0.671	0.139	0.139	2
Jaya Santosa	2	0.669	0.084	0.084	2
Gallery Cat / CV. Cahaya Bintang Sejati	2	0.959	0.535	0.535	2
Karya Subur	1	0.204	0.795	0.204	1
Maju	2	0.544	0.161	0.161	2
Putra Jaya	1	0.277	0.360	0.277	1
Samudra Jaya	2	0.676	0.144	0.144	2
Satria	2	0.684	0.131	0.131	2
Sejahtera	2	0.658	0.109	0.109	2
Sinar Jaya II	2	0.536	0.190	0.190	2
Sinar Langgeng	2	0.471	0.201	0.201	2
Subur II	1	0.120	0.715	0.120	1
Surya	1	0.014	0.586	0.014	1
Wisma Harapan	2	1.336	0.992	0.992	2
Kharisma Jaya	2	0.437	0.229	0.229	2
Sumber Rejeki	2	1.270	0.936	0.936	2
CV. Sejahtera	2	0.556	0.080	0.080	2

<i>Retailer</i>	K- Means Cluster	Jarak ke Centroid		Min	Hasil
		1	2		
Sumber Rejeki II	2	1.150	0.827	0.827	2
CV.Seng Joyo Mandiri	2	0.551	0.107	0.107	2
Makmur Abadi	2	0.933	0.523	0.523	2
J O Y O	2	1.123	0.935	0.935	2
Budhi Jaya	2	0.458	0.190	0.190	2
Dwi Tunggal	2	0.448	0.216	0.216	2
Jaya Abadi	1	0.114	0.491	0.114	1
Karunia Jaya	2	0.642	0.152	0.152	2
Naga Jaya	1	0.252	0.363	0.252	1
Naga Emas	2	0.415	0.223	0.223	2
Putra Makmur	2	0.491	0.169	0.169	2
Putra Mas	2	0.544	0.166	0.166	2
Sinar Abadi	2	0.973	0.561	0.561	2
Sinar Agung	2	0.670	0.107	0.107	2
Warna	2	0.552	0.173	0.173	2
Kota Jaya	2	0.547	0.181	0.181	2
Pastelle Interior	2	0.852	0.409	0.409	2
Dua Jaya Tunggal	2	0.734	0.222	0.222	2

C-8

<i>Retailer</i>	K-Means Cluster	Jarak ke Centroid		Min	Hasil
		1	2		
Surya Makmur	2	0.604	0.150	0.150	2
Agung Makmur	2	0.735	0.222	0.222	2
G M (UD. Hm. Masri S.)	1	0.011	0.605	0.011	1
Saputra Jaya	2	0.664	0.120	0.120	2
Aneka Logam	2	0.312	0.303	0.303	2
Anugrah	1	0.300	0.325	0.300	1
Baru	1	0.115	0.486	0.115	1
Gangsar	2	0.610	0.131	0.131	2
Pacific Raya Motor	1	0.053	0.546	0.053	1
Sinar Indah	2	0.505	0.189	0.189	2
Sumber Utama Karya	2	0.656	0.119	0.119	2
Wahyu	2	0.650	0.116	0.116	2
Sunrise	2	0.443	0.210	0.210	2
Warna Jaya	2	0.544	0.166	0.166	2
Subur	2	0.550	0.159	0.159	2
Ultimate Shine	2	0.551	0.103	0.103	2
Proline	1	0.240	0.832	0.240	1
Bandar Jaya	1	0.254	0.846	0.254	1

<i>Retailer</i>	K- Means Cluster	Jarak ke Centroid		Min	Hasil
		1	2		
Habza	2	0.527	0.125	0.125	2
Guna Jaya	1	0.159	0.753	0.159	1
Pratama	1	0.114	0.488	0.114	1
Mukti Jaya	1	0.258	0.357	0.258	1
Makmur Jaya	1	0.078	0.674	0.078	1
Mahkota Auto Paint	2	0.367	0.277	0.277	2
Leo	2	0.544	0.169	0.169	2
Aa Elektro	2	0.703	0.188	0.188	2
Bangun Rejeki	2	0.530	0.160	0.160	2
Istana	1	0.204	0.797	0.204	1
Sinar Jaya	2	0.651	0.167	0.167	2
Lidah Utomo	2	0.663	0.121	0.121	2
Tri Stars	2	0.392	0.253	0.253	2
Subur Jaya III	1	0.018	0.591	0.018	1
Artomoro	2	0.719	0.174	0.174	2
Valao	2	0.517	0.099	0.099	2
Cahaya Makmur	2	0.506	0.159	0.159	2
Hartono Jaya	2	0.677	0.136	0.136	2

C-10

<i>Retailer</i>	K-Means Cluster	Jarak ke Centroid		Min	Hasil
		1	2		
Yudi	2	0.590	0.113	0.113	2
Prado	2	0.724	0.178	0.178	2
Evan Jaya Makmur	2	0.715	0.155	0.155	2
Adi Jaya	2	0.681	0.113	0.113	2
A.A	1	0.255	0.367	0.255	1
Sumber Rejeki III	2	0.651	0.166	0.166	2
Teng Joyo	2	0.672	0.100	0.100	2
Inti Varian	1	0.176	0.769	0.176	1
UD. Rukun Jaya	1	0.330	0.920	0.330	1
CV. Beruang Sukses Makmur	1	0.073	0.530	0.073	1
Varian	2	0.552	0.179	0.179	2
Insan Jaya	2	0.449	0.223	0.223	2

Lampiran D KUISIONER AHP



KUISIONER SURVEY PERINGKATAN KEPENTINGAN KRITERIA PADA PERILAKU RETAILER PT. BINA ADIDAYA SURABAYA DALAM MELAKUKAN TRANSAKSI



Kuisisioner ini bertujuan untuk mendapatkan masukan pendapat/pemikiran dari Narasumber terkait dengan peringkat kriteria yang akan digunakan dalam pemberian bobot pada variabel segmentasi pelanggan. Kami menjamin kerahasiaan data/informasi yang diberikan serta tidak akan membawa implikasi apapun bagi responden.

IDENTITAS RESPONDEN

Nama Responden : _____

Jabatan : _____

Lama Bekerja : _____ Tahun

PETUNJUK PENGISIAN KUISIONER

1. Berilah tanda checklist (√) untuk menunjukkan persepsi atau penilaian bapak/ibu terhadap kriteria *retailer* bagaimanakah yang dipilih lebih penting dengan melihat perilaku *retailer* dalam melakukan transaksi dengan perusahaan. (Bandingkan kriteria yang ada pada kolom sebelah kiri dengan kriteria yang ada pada kolom sebelah kanan)

Skala numerik akan menunjukkan suatu perbandingan dari tingkat kepentingan dua kriteria dengan penjelasan setiap skalanya yaitu:

D-2

Skala Nilai	Tingkat Preferensi	Penjelasan
1	Sama pentingnya	Kedua kriteria memiliki pengaruh yang sama
3	Sedikit lebih penting	Penilaian dan pengalaman sedikit memihak pada salah satu kriteria tertentu dibandingkan kriteria pasangannya.
5	Lebih penting	Penilaian dan pengalaman memihak pada salah satu kriteria tertentu dibandingkan kriteria pasangannya.
7	Jelas lebih penting	Salah satu kriteria lebih diprioritaskan dan relatif lebih penting dibandingkan kriteria pasangannya.
9	Mutlak sangat penting	Salah satu kriteria sangat jelas lebih penting dibandingkan kriteria pasangannya.
2, 4, 6, 8		Diberikan bila terdapat keraguan penilaian di antara dua tingkat kepentingan yang berdekatan.

2. Jika kriteria pada sebelah kiri lebih penting dari pada kriteria sebelah kanan maka, pilih skala yang dekat dengan sebelah kiri dan jika sebaliknya pilih skala yang dekat dengan sebelah kanan.
3. Kriteria pembobotan terhadap *retailer* terdapat tiga macam yaitu:
 - *Recency* : jarak antara waktu terakhir transaksi dengan waktu saat ini.
 - *Frequency* : seberapa sering jumlah transaksi yang dilakukan oleh *retailer* pada periode tertentu.

Monetary : jumlah uang yang dihabiskan *retailer* saat transaksi pada periode tertentu.

CONTOH PENGISIAN KUISIONER

Kriteria																		Kriteria
Recency	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Frequency
															√			

Artinya:

Pada pernyataan tersebut, tanda checklist (√) di kolom jawaban sebelah kiri menunjukkan bahwa: “*Recency*” $1/7$ kali lebih penting dibandingkan dengan “*Frequency*” dalam hal peringkat kriteria untuk *retailer*. Dengan kata lain “*Frequency*” 7 kali lebih penting dibandingkan dengan “*Recency*” dalam hal peringkat kriteria untuk *retailer*.

DAFTAR PERTANYAAN

Berikan tanda checklist (√) untuk menunjukkan persepsi atau penilaian bapak/ibu terhadap perbandingan kriteria *retailer* yang dipilih lebih penting dengan melihat perilaku *retailer* dalam melakukan transaksi dengan perusahaan. “Manakah kriteria yang lebih penting antara kriteria pada kolom sebelah kiri dengan kriteria pada kolom sebelah kanan?”

Kriteria																		Kriteria
Recency	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Frequency
Recency	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Monetary
Frequency	9	8	7	6	5	4	3	2	1	2	3	4	5	6	7	8	9	Monetary

***-Terima kasih atas waktu yang telah diberikan Bapak/Ibu
untuk mengisi kuisisioner ini-***

D-4

Halaman ini sengaja dikosongkan

Lampiran E USE CASE DESCRIPTION

Tabel E.1 Use case description “Melihat persebaran *retailer* pada grafik 3 dimensi”

UC-2 Melihat persebaran <i>retailer</i> pada grafik 3 dimensi		
Purpose	Mengetahui persebaran setiap <i>retailer</i> berdasarkan <i>recency</i> , <i>frequency</i> dan <i>monetary</i> pada grafik 3 dimensi	
Overview	Dimulai ketika data hasil proses <i>clustering</i> telah dimasukkan pada script aplikasi	
Actors	Pengguna aplikasi	
Pre Condition	Pengguna telah membuka aplikasi	
Past Condition	Grafik “3 dimensi”	
Typical Course Event	Actor	System
	1. Klik icon slide bar  3. Klik menu detail 2 segmen atau detail 3 segmen 5. Klik salah satu jenis segmen 7. Klik dan geser salah satu sisi grafik	2. Menampilkan slide bar pilihan menu 4. Menampilkan halaman detail 2 atau 3 segmen 6. Menghasilkan grafik 3 dimensi 8. Menghasilkan grafik dari sisi yang berbeda
Alternate Flow of Events	Step 2: tidak mengklik menu detail 2 segmen atau detail 3 segmen, maka aplikasi akan memilih menu default yaitu 2 segmen Step 5: Tidak mengklik salah satu jenis segmen, maka aplikasi akan memilih segmen default yaitu semua segmen	
Exceptional Flow of Event	Aplikasi tertutup otomatis karena error	

Tabel E.2 Use case description “Melihat persebaran *retailer* pada peta lokasi”

UC-3 Melihat persebaran <i>retailer</i> pada peta lokasi		
Purpose	Mengetahui persebaran segmen setiap <i>retailer</i> pada peta lokasi	
Overview	Dimulai ketika data hasil proses <i>clustering</i> telah dimasukkan pada script aplikasi	
Actors	Pengguna aplikasi	
Pre Condition	Pengguna telah membuka aplikasi	
Past Condition	Peta Lokasi	
Typical Course Event	Actor	System
	1. Klik icon slide bar  3. Klik menu detail 2 segmen atau detail 3 segmen 5. Klik tab menu peta lokasi 7. Klik salah satu jenis segmen 9. Klik titik <i>retailer</i> pada peta	2. Menampilkan slide bar pilihan menu 4. Menampilkan halaman detail 2 atau 3 segmen 6. Menghasilkan peta lokasi 9. Menghasilkan titik <i>retailer</i> yang termasuk dalam segmen yang pilih 10. Menampilkan popup informasi nama <i>retailer</i> dan nilai RFM
Alternate Flow of Events	Step 2: tidak mengklik menu detail 2 segmen atau detail 3 segmen, maka aplikasi akan memilih menu default yaitu 2 segmen Step 7: Tidak mengklik salah satu jenis segmen, maka aplikasi akan memilih segmen default yaitu semua segmen	
Exceptional Flow of Event	Aplikasi tertutup otomatis karena error	

Tabel E.3 Use case description “Melihat peringkat *retailer*”

UC-4 Melihat peringkat <i>retailer</i>		
Purpose	Mengetahui <i>retailer</i> terbaik dan nilai CLV setiap <i>retailer</i>	
Overview	Dimulai ketika data hasil proses <i>clustering</i> telah dimasukkan pada script aplikasi	
Actors	Pengguna aplikasi	
Pre Condition	Pengguna telah membuka aplikasi	
Past Condition	<i>retailer</i> terbaik dan terburuk serta tabel data nilai CLV	
Typical Course Event	Actor	System
	<ol style="list-style-type: none"> 1. Klik icon slide bar  3. Klik menu detail 2 segmen atau detail 3 segmen 5. Klik icon plus  7. Klik tab menu peringkat <i>retailer</i> 9. Klik salah satu jenis segmen 	<ol style="list-style-type: none"> 2. Menampilkan slide bar pilihan menu 4. Menampilkan halaman detail 2 atau 3 segmen 6. Menghasilkan nama <i>retailer</i> terbaik dan terburuk 8. Menampilkan tabel data nilai CLV 10. Menampilkan nama <i>retailer</i> terbaik dan baris data sesuai jenis segmen yang dipilih
Alternate Flow of Events	<p>Step 2: tidak mengklik menu detail 2 segmen atau detail 3 segmen, maka aplikasi akan memilih menu default yaitu 2 segmen</p> <p>Step 9: Tidak mengklik salah satu jenis segmen, maka aplikasi akan memilih segmen default yaitu semua segmen</p>	
Exceprional Flow of Event	Aplikasi tertutup otomatis karena error	

Tabel E.4 Use case description “Melihat persebaran variabel setiap segmen pada box plot”

UC-5 Melihat persebaran variabel setiap segmen pada box plot		
Purpose	Mengetahui persebaran <i>recency</i> , <i>frequency</i> dan <i>monetary</i> setiap segmen pada box plot	
Overview	Dimulai ketika data hasil proses <i>clustering</i> telah dimasukkan pada script aplikasi	
Actors	Pengguna aplikasi	
Pre Condition	Pengguna telah membuka aplikasi	
Past Condition	Box Plot setiap variabel	
Typical Course Event	Actor	System
	1. Klik icon slide bar  3. Klik menu detail 2 segmen atau detail 3 segmen 5. Klik icon plus 	2. Menampilkan slide bar pilihan menu 4. Menampilkan halaman detail 2 atau 3 segmen 6. Menghasilkan box plot setiap variabel
Alternate Flow of Events	Step 2: tidak mengklik menu detail 2 segmen atau detail 3 segmen, maka aplikasi akan memilih menu default yaitu 2 segmen	
Exceptional Flow of Event	Aplikasi tertutup otomatis karena error	

Tabel E.5 Use case description “Melihat range setiap segmen”

UC-6 Melihat range setiap segmen		
Purpose	Mengetahui range berupa nilai maksimal dan minimal <i>recency</i> , <i>frequency</i> dan <i>monetary</i> di setiap segmen	
Overview	Dimulai ketika data hasil proses <i>clustering</i> telah dimasukkan pada script aplikasi	
Actors	Pengguna aplikasi	
Pre Condition	Pengguna telah membuka aplikasi	
Past Condition	Range segmen	
Typical Course Event	Actor	System
	1. Klik icon slide bar  3. Klik menu detail 2 segmen atau detail 3 segmen 5. Klik salah satu jenis segmen	2. Menampilkan slide bar pilihan menu 4. Menampilkan halaman detail 2 atau 3 segmen 8. Menghasilkan range segmen
Alternate Flow of Events	Step 2: tidak mengklik menu detail 2 segmen atau detail 3 segmen, maka aplikasi akan memilih menu default yaitu 2 segmen Step 5: Tidak mengklik salah satu jenis segmen, maka aplikasi akan memilih segmen default yaitu semua segmen	
Exceprional Flow of Event	Aplikasi tertutup otomatis karena error	

Tabel E.6 Use case description “Melihat rata-rata nilai variabel setiap segmen”

UC-7 Melihat rata-rata nilai variabel setiap segmen		
Purpose	Mengetahui rata-rata <i>recency</i> , <i>frequency</i> dan <i>monetary</i> setiap segmen	
Overview	Dimulai ketika data hasil proses <i>clustering</i> telah dimasukkan pada script aplikasi	
Actors	Pengguna aplikasi	
Pre Condition	Pengguna telah membuka aplikasi	
Past Condition	Rata-rata <i>recency</i> , <i>frequency</i> dan <i>monetary</i>	
Typical Course Event	Actor	System
	1. Klik icon slide bar  3. Klik menu detail 2 segmen atau detail 3 segmen 5. Klik salah satu jenis segmen	2. Menampilkan slide bar pilihan menu 4. Menampilkan halaman detail 2 atau 3 segmen 6. Menghasilkan rata-rata <i>recency</i> , <i>frequency</i> dan <i>monetary</i>
Alternate Flow of Events	Step 2: tidak mengklik menu detail 2 segmen atau detail 3 segmen, maka aplikasi akan memilih menu default yaitu 2 segmen Step 5: Tidak mengklik salah satu jenis segmen, maka aplikasi akan memilih segmen default yaitu semua segmen	
Exceptional Flow of Event	Aplikasi tertutup otomatis karena error	

Tabel E.7 Use case description “Melihar data tabel”

UC-8 Melihat tabel data		
Purpose	Mengetahui data secara keseluruhan dalam sebuah tabel	
Overview	Dimulai ketika data hasil proses <i>clustering</i> telah dimasukkan pada script aplikasi	
Actors	Pengguna aplikasi	
Pre Condition	Pengguna telah membuka aplikasi	
Past Condition	Baris data setiap <i>retailer</i>	
Typical Course Event	Actor	System
	1. Klik icon slide bar  3. Klik menu tabel data 5. Memasukkan keyword pencarian 7. Memilih banyak baris yang ditampilkan 9. Klik halaman tabel yang diinginkan	2. Menampilkan slide bar pilihan menu 4. Menghasilkan tabel data secara keseluruhan 6. Menampilkan baris data sesuai keyword 8. Menampilkan baris data sesuai dengan pilihan 10. Menampilkan halaman tabel yang dipilih
Alternate Flow of Events	Step 2: tidak mengklik menu tabel data, maka aplikasi akan memilih menu default yaitu 2 segmen	
Exceptional Flow of Event	Aplikasi tertutup otomatis karena error	

Halaman ini sengaja dikosongkan

Lampiran F

KODE PROGRAM *SHINY*

Kode program Server.R untuk 3 segmen

```
library(shiny)
library(leaflet)
library(plot3Drgl)
library(sqldf)
library(shinydashboard)
library(plotly)

function(input, output, session) {
  #ALL
  alldata <- read.csv("D:/TA/DASHBOARD/all.csv", sep=";")
  ta.data <- read.csv("D:/TA/DASHBOARD/adidaya.csv", sep=";")
  pilihdata <- reactive({
    switch(input$daset,
           "Semua Segmen" = ta.data,
           "Segmen 1" = subset(ta.data, ta.data$klasterke==1),
           "Segmen 2" = subset(ta.data, ta.data$klasterke==2),
           "Segmen 3" = subset(ta.data, ta.data$klasterke==3)
    )
  })
  #Info
  output$ket1<- renderValueBox({
    valueBox(
      value = "Surabaya",
      subtitle = "(Area Pelanggan)",
      icon = icon("map-marker"),
      color = "teal"
    )
  })
  #Info
  output$ket2<- renderValueBox({
    valueBox(
      value = "Jan s/d Des 2016",
      subtitle = "(Data Transaksi)",
      icon = icon("calendar"),
      color = "blue"
    )
  })
  #Info
  output$ket3<- renderValueBox({
    valueBox(
      value = "Pemasaran Retail",
      subtitle = "(Kategori Produk)",
      icon = icon("shopping-cart"),
      color = "purple"
    )
  })
  #Data Bar dan Pie
  segmen <- c("Segmen 1", "Segmen 2", "Segmen 3")
  value <- c(sum(ta.data$klasterke == "1"),
            sum(ta.data$klasterke == "2"),
            sum(ta.data$klasterke == "3"))
  #Warna Cluster
  warna <- c("#D2691E", "#1E90FF", "#DC143C")
}
```

```

#BAR
output$bar <- renderPlotly ({
  plot_ly(
    x = segmen,
    y = value,
    name = "Jenis Segmen",
    type = "bar",
    hoverinfo = 'text',
    text = paste(value, ' Retail'),
    marker = list(color = warna)
  )
})

#PIE
output$pie <- renderPlotly ({
  plot_ly(labels = segmen, values = value, type = 'pie',
    textposition = 'inside',
    textinfo = 'label+percent',
    insidetextfont = list(color = '#FFFFFF'),
    hoverinfo = 'text',
    text = paste(value, ' Retail'),
    marker = list(colors = warna,
    line = list(color = '#FFFFFF', widt
h = 1))
  )
})

#Judul Pilihan
output$judul <- renderText({
  paste("Menampilkan Detail",input$dataset)
})

#Rata-rata Recency
output$recency <- renderInfoBox({
  data <- pilihdata()
  infoBox(
    title = "Recency",
    value = paste(round(mean(data$recency)) ,"hari"),
    subtitle = "Terakhir transaksi",
    icon = icon("calendar"),
    color = "teal",
    fill = TRUE
  )
})

#Rata-rata Frequency
output$frequency <- renderInfoBox({
  data <- pilihdata()
  infoBox(
    title = "Frequency",
    value = paste(round(mean(data$frequency)) ,"kali"),
    subtitle = "Banyak transaksi",
    icon = icon("area-chart"),
    color = "teal",
    fill = TRUE
  )
})

```

```

#Rata-rata Monetary
output$monetary <- renderInfoBox({
  data <- pilihdata()
  infoBox(
    title = "Monetary",
    value = paste("Rp", round(mean(data$monetary))),
    subtitle = "jumlah biaya",
    icon = icon("money"),
    color = "teal",
    fill = TRUE
  )
})

#Retailer Terbaik
output$baik<- rendervalueBox({
  re <- pilihdata()
  cari <- sqldf("select retailer From
                (select retailer,max(clv) From re)")
  valueBox(
    value = tags$h5( "RETAILER TERBAIK"),
    subtitle = tags$h4(tags$strong (cari)),
    icon = icon("thumbs-o-up"),
    color = "purple"
  )
})

#Retailer Terburuk
output$buruk<- rendervalueBox({
  re <- pilihdata()
  cari <- sqldf("select retailer From (select retailer,min(
clv) From re)")
  valueBox(
    value = tags$h5( "RETAILER TERBURUK"),
    subtitle = tags$h4(tags$strong (cari)),
    icon = icon("thumbs-o-down"),
    color = "purple"
  )
})

#Tulisan Rentang Nilai
output$text <- renderText({
  paste("Rentang Nilai", input$dataset)
})

#rentang Nilai RFM
output$range <- renderPrint({
  rentang <- pilihdata()
  list(Recency = paste(min(rentang$recency), "-",
                        max(rentang$recency), "hari"),
       Frequency = paste (min(rentang$frequency), "-",
                           max(rentang$frequency), "kali"),
       Monetary = paste ("Rp",min(rentang$monetary), "-",
                          "Rp",max(rentang$monetary))
  )
})

#Data box plot
y1 <- subset(ta.data,ta.data$klasterke==1)
y2 <- subset(ta.data,ta.data$klasterke==2)
y3 <- subset(ta.data,ta.data$klasterke==3)

```

```

#Box Plot
output$boxr <- renderPlotly ({
  plot_ly(type = 'box') %>%
    add_boxplot(y = y1$recency, name = "1",
               marker = list(color = "#D2691E"),
               line = list(color = "#D2691E")) %>%
    add_boxplot(y = y2$recency, name = "2",
               marker = list(color = "#1E90FF"),
               line = list(color = "#1E90FF")) %>%
    add_boxplot(y = y3$recency, name = "3",
               marker = list(color = "#DC143C"),
               line = list(color = "#DC143C"))
})

output$boxf <- renderPlotly ({
  plot_ly(type = 'box') %>%
    add_boxplot(y = y1$frequency, name = "1",
               marker = list(color = "#D2691E"),
               line = list(color = "#D2691E")) %>%
    add_boxplot(y = y2$frequency, name = "2",
               marker = list(color = "#1E90FF"),
               line = list(color = "#1E90FF")) %>%
    add_boxplot(y = y3$frequency, name = "3",
               marker = list(color = "#DC143C"),
               line = list(color = "#DC143C"))
})

output$boxm <- renderPlotly ({
  plot_ly(type = 'box') %>%
    add_boxplot(y = y1$monetary, name = "1",
               marker = list(color = "#D2691E"),
               line = list(color = "#D2691E")) %>%
    add_boxplot(y = y2$monetary, name = "2",
               marker = list(color = "#1E90FF"),
               line = list(color = "#1E90FF")) %>%
    add_boxplot(y = y3$monetary, name = "3",
               marker = list(color = "#DC143C"),
               line = list(color = "#DC143C"))
})

#Grafik 3dimensi
output$thewidget1 <- renderRglwidget ({
  so <- pilihdata()
  open3d()
  plot3d(x=so$recency, y=so$frequency, z=so$monetary,
        col = warna[so$klasterke],
        type = 's', size = 1,
        xlab = "recency",
        ylab = "frequency",
        zlab = "monetary",
        colkey = list(length = 0.5, width = 0.5,
                      cex.clab = 0.75)
  )
  scene1 <- scene3d()
  rgl.close()

  save <- options(rgl.inShiny = TRUE)
  on.exit(options(save))

  rglwidget(scene1)
})

```

```

#Peta Lokasi
output$map<- renderLeaflet({
  dat <- pilihdata()
  leaflet() %>%
    addTiles() %>%
    setView(112.73439800,-7.28916600, zoom = 12) %>%
    addCircleMarkers(data = dat, ~lng, ~lat,
                      radius = 3,
                      color = warna[dat$klasterke],
                      popup = paste(
                        "<b>", dat$retailer,"</b>","<br>",
                        "Recency : ", dat$recency,"hari", "
<br>",
                        "Frequency : ", dat$frequency,"kali"
                        "Monetary : ", "Rp", dat$monetary
                      )
    )
})

#CLV
output$clv <- renderDataTable({
  tabel <- pilihdata()
  kluaran <- data.frame(tabel$code,
                        tabel$retailer,
                        tabel$clv
  )
  colnames(kluaran) <- c("Code",
                        "Retailer",
                        "CLV")
  print(kluaran)
});
list(lengthMenu = c(5, 10, 15), pageLength = 8))

#Keterangan warna
output$tabelsegmen <- renderUI({
  dircolors <-c("1"="#D2691E", "2"="#1E90FF", "3"="#DC143C"
)
  tags$table(class = "table",
             tags$thead(tags$str(
               tags$th("Color"),
               tags$th("Segmen")
             )),
             tags$body(
               tags$str(
                 tags$td(span(style = sprintf(
                   "width:1.1em; height:1.1em; background-c
olor:%s; display:
                   inline-block;",
                   dirColors[1]
                 )),
                 tags$td("Segmen 1")
               )
             )
)

```

```

        tags$str(
          tags$td(span(style = sprintf(
            "width:1.1em; height:1.1em; background-c
olor:%s; display:
            inline-block;",
            dirColors[2]
          )),
          tags$td("segmen 2")
        ),
        tags$str(
          tags$td(span(style = sprintf(
            "width:1.1em; height:1.1em; background-c
olor:%s; display:
            inline-block;",
            dirColors[3]
          )),
          tags$td("segmen 3")
        )
      )
    }
  })
}

#Data Tabel
output$mytable = renderDataTable({
  tabel <- alldata
  kluaran <- data.frame(tabel$code,
    tabel$retailer,
    tabel$recency,
    tabel$frequency,
    tabel$monetary,
    tabel$clv,
    tabel$dua,
    tabel$tiga
  )
  colnames(kluaran) <- c("Code",
    "Retailer",
    "Recency",
    "Frequency",
    "Monetary",
    "CLV",
    "2 Segmen",
    "3 Segmen"
  )
  print(kluaran)
})
list(lengthMenu = c(5, 10, 15), pageLength = 10))
}

```

Kode program Ui.R untuk 3 segmen

```

library(shiny)
library(leaflet)
library(plot3Drgl)
library(sqldf)
library(shinydashboard)
library(plotly)

dashboardPage(skin = "black",
  dashboardHeader(title = "PT. Bina Adidaya"),
  dashboardSidebar(
    sidebarMenu(
      menuItem("2 Segmen", tabName = "home2", icon = icon("home")),
      menuItem("Detail 2 Segmen", tabName = "detail2", icon = icon("dashboard")),
      menuItem("3 Segmen", tabName = "home", icon = icon("home")),
      menuItem("Detail 3 Segmen", tabName = "detail", icon = icon("dashboard")),
      menuItem("Tabel Data", tabName = "tabel", icon = icon("table"))
    )
  ),
  dashboardBody(
    tabItems(
      tabItem("home",
        fluidRow(
          column(width = 2,
            h1("Hasil Segmentasi Pelanggan PT. Bina Adidaya"),
            p("sebanyak", strong("83"), strong(em("retailer"))),
            p("di Surabaya dapat terbagi menjadi", strong("3 segmen.")),
            p("Nama segmen merupakan urutan dari segmen terbaik hingga segmen terburuk.")
          ),
          column(width = 5,
            box(title = "Jumlah Anggota Segmen",
              status = "info",
              width = 50,
              plotlyOutput("bar"),
              solidHeader = TRUE
            )
          ),
          column(width = 5,
            box(title = "Persentase Jumlah Anggota Segmen",
              status = "info",
              width = 50,
              plotlyOutput("pie"),
              solidHeader = TRUE
            )
          )
        )
      )
    )
  )
)

```

```

        fluidRow(
          valueBoxOutput("ket1"),
          valueBoxOutput("ket2"),
          valueBoxOutput("ket3")
        ),
      ),
      tabItem("detail",
        fluidPage(
          column(width=3,
            box(width = NULL, status = "warning",
              background = "light-blue",
              h4(strong(textOutput('judul'))),
              br(),
              selectInput("dataset", "Pilih Jenis
Segmen",
                                choices = c(
                                  "Semua Segmen",
                                  "Segmen 1",
                                  "Segmen 2",
                                  "Segmen 3")
                                )
            ),
            box(width = 330, background = "aqua",
              strong(textOutput('text')),
              verbatimTextOutput('range')
            ),
            box(width = NULL,
              uiOutput("tabelsegmen")
            )
          ),
          column(width=9,
            fluidRow(
              infoBoxOutput("recency"),
              infoBoxOutput("frequency"),
              infoBoxOutput("monetary")
            ),
            fluidRow(
              box( width = 100,
                title = "Peringkat Retailer",
                background = "purple",
                solidHeader = TRUE,
                collapsible = TRUE,
                collapsed = TRUE,
                valueBoxOutput("baik"),
                valueBoxOutput("buruk")
              )
            ),
          ),

```


F-10

Halaman ini sengaja dikosongkan