



**SEGMENTASI CLV PELANGGAN RETAIL MENGGUNAKAN MODEL
LRFM DAN METODE AHC UNTUK MENENTUKAN STRATEGI CRM**

SKRIPSI

Oleh:

Septia Puji Lestari

NIM 182410103019

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS JEMBER

2023



**SEGMENTASI CLV PELANGGAN RETAIL MENGGUNAKAN
MODEL LRFM DAN METODE AHC UNTUK MENENTUKAN
STRATEGI CRM**

SKRIPSI

Diajukan guna melengkapi tugas akhir dan memenuhi salah satu syarat untuk menyelesaikan Pendidikan Sarjana (S1) Program Studi Informatika Universitas Jember dan mencapai gelar Sarjana Komputer

Oleh:

Septia Puji Lestari

NIM 182410103019

PROGRAM STUDI INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS JEMBER

2023

PERSEMBAHAN

Skripsi ini saya persembahkan untuk:

1. Allah SWT yang senantiasa memberikan rahmat dan hidayah-Nya yang senantiasa memberikan kekuatan, kesabaran, kemudahan dan kelancaran dalam mengerjakan skripsi.
2. Bapak Sariyanto dan Ibu Endang Lestari, atas segala perjuangan, pengorbanan, kesabaran, keikhlasan, do'a dan kasih sayang yang tidak terukur lagi oleh apapun selama hidup saya.
3. Pariera Vanesa Purbasari adik saya tercinta yang senantiasa mendengarkan keluh kesah saya dan menjadi pelengkap hidup saya.
4. Teman-teman SMA, Kuliah, Angkatan DECISION 18 yang telah menemani dan menjadi teman seperjuangan selama menempuh pahitnya hidup, manisnya rasa keberhasilan selama menempuh pendidikan di Universitas Jember

MOTTO

“Sebaik-baiknya rencanamu masih baik rencana Tuhanmu”

“Get your diploma”

(Choi Seung Cheol)



PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Septia Puji Lestari

NIM : 182410103019

menyatakan dengan sesungguhnya bahwa karya ilmiah dengan judul “Segmentasi CLV Pelanggan Retail Menggunakan Model LRFM Dan Metode AHC Untuk Menentukan Strategi CRM adalah benar – benar hasil karya sendiri, kecuali jika dalam kutipan substansi disebutkan sumbernya, belum pernah diajukan pada institusi mana pun, dan bukan merupakan karya jiplakan. Saya bertanggung jawab atas keabsahan dan kebenaran isinya sesuai dengan sikap ilmiah yang harus dijunjung tinggi.

Demikian pernyataan ini saya buat dengan sebenarnya, tanpa adanya tekanan dan paksaan dari pihak manapun serta bersedia mendapat sanksi akademik jika di kemudian hari pernyataan ini tidak benar.

Jember, 26 Januari 2022

Yang menyatakan

Septia Puji Lestari

NIM 182410103019

PENGESAHAN PEMBIMBING

Skripsi berjudul “Segmentasi CLV Pelanggan Retail Menggunakan Model LRFM dan Metode AHC Untuk Menentukan Strategi CRM”, telah diuji dan disahkan pada:

Hari,tanggal : Rabu, 26 Januari 2023

Tempat : Fakultas Ilmu Komputer, Universitas Jember.

Disetujui oleh:

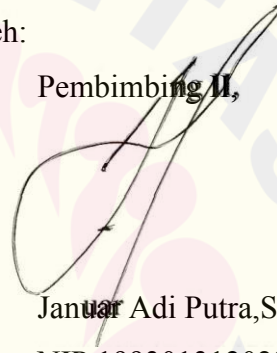
Pembimbing I,



Prof. Dr.Saiful Bukhori,ST.,M.Kom.

NIP. 196811131994121001

Pembimbing II,



Januar Adi Putra,S.Kom.,M.Kom

NIP.199301312022031005

PENGESAHAN PENGUJI

Skripsi berjudul “Segmentasi CLV Pelanggan Retail Menggunakan Model LRFM dan Metode AHC Untuk Menentukan Strategi CRM”, telah diuji dan disahkan pada:

Hari,tanggal : Rabu, 26 Januari 2023

Tempat : Fakultas Ilmu Komputer, Universitas Jember.

Disetujui oleh:

Penguji I,



Nelly Oktavia Adiwijaya S.Si.,MT.

NIP. 198410242009122008

Penguji II,



Muhamad Arief Hidayat S.Kom.,M.Kom.

NIP. 198101232010121003

Mengesahkan

Dekan Fakultas Ilmu Komputer,



Drs. Antonius Cahya Prihandoko, M. App.Sc., ph.D.

NIP. 196909281993021001

RINGKASAN

Segmentasi CLV Pelanggan Retail Menggunakan Model LRFM Dan Metode AHC untuk Menentukan Strategi CRM; Septia Puji Lestari; 182410103019; 2023; 79 halaman; Prodi Informatika Fakultas Ilmu Komputer Universitas Jember.

Retail modern seperti gerai toko swalayan atau supermarket mengalami peningkatan permintaan (Gunawan 2022). Menurut data dari Bank Indonesia, Penjualan ecer pada bulan April 2022 diperkirakan akan meningkat secara bulanan. Hal ini terlihat dari Indeks Penjualan (IPR) April tahun 2022 yang tercatat sebesar 219,3 atau secara bulanan penjualan tumbuh ke angka 6,8% (Haryono 2022). Peningkatan penjualan perlu dilakukan untuk mencegah terjadinya penurunan penjualan. Pencegahan tersebut dapat dilakukan dengan memperhatikan pelanggan secara teliti yang berguna untuk mengurangi kehilangan pelanggan (Rahmawati, Toana, dan Sofyan 2021). Selain itu, toko retail juga harus mempertahankan pelanggan yang memiliki nilai tinggi terhadap perusahaan (Marisa dkk. 2019).

CRM merupakan strategi bisnis untuk mendapatkan keuntungan, pendapatan, dan kepuasan dari pelanggan serta bagaimana cara memperlakukan pelanggan (Justitia, Hidayat, dan Santoso 2021). Penyusunan strategi penjualan dapat dilakukan dengan melakukan analisa CRM. Analisa CRM ini berkaitan dengan membagi pelanggan dengan sifat dan kriteria sama, ini dilakukan untuk mengetahui minat belanja masyarakat terhadap toko dan pelanggan yang memiliki loyalitas terhadap toko. Untuk mengetahui loyalitas dapat dilakukan dengan menghitung nilai CLV. Model yang digunakan untuk mendapatkan nilai CLV adalah LRFM. Model LRFM merupakan perkembangan dari model RFM dengan penambahan variabel length untuk membedakan pelanggan yang memiliki hubungan jangka pendek maupun jangka Panjang dengan perusahaan.

Segmentasi pelanggan dapat dilakukan dengan menggunakan Machine learning clustering. Metode yang digunakan adalah *Agglomerative Hierarchical Clustering* Algoritma *Agglomerative Hierarchical Clustering* merupakan algoritma yang digunakan untuk mengelompokkan data dengan mengasumsikan jumlah cluster terlebih dahulu namun pengelompokkan berdasarkan dengan kesamaan

antar jarak atau dissimilarity. Konsep algoritma ini menggambarkan semua data menjadi satu cluster kecil (leaf) lalu dikelompokkan dengan cluster lain dan membentuk cluster besar (nodes). Metode Agglomerative menghitung nilai (dis)similarity dengan menggunakan panjang garis antar data (garis tegak lurus) pada pohon dendogram.

Tujuan dari penelitian ini adalah mengsegmentasi customer lifetime value pelanggan retail berdasarkan model LRFM dengan menggunakan metode *Agglomerative Hierarchical Clustering*. Serta untuk menentukan strategi CRM yang dapat diberikan pada toko retail berdasarkan hasil segmentasi.

Hasil dari proses clustering pada data transaksi penjualan KPRI UNEJ. Didapatkan 2 cluster pelanggan. dimana cluster 0 merupakan pelanggan kategori *new customer* dengan jumlah anggota cluster sebanyak 998 pelanggan. Nilai CLV untuk pelanggan tersebut adalah 0.288768 lebih rendah dari cluster 1 dimana nilai cluster 1 merupakan pelanggan kategori *potensial customer* dengan jumlah anggota cluster sebanyak 1005 pelanggan. Nilai CLV untuk pelanggan tersebut adalah 0.336871 lebih tinggi dari cluster 0.

Strategi CRM yang dapat dilakukan oleh perusahaan untuk kategori *new customer* menurut (Alizadeh Zoeram dan Karimi Mazidi 2018) terkait pelanggan dengan kategori new customer disarankan perusahaan untuk menaruh lebih banyak perhatian. Salah satu contohnya adalah memberi transactional satisfaction kepada pelanggan dengan kategori new customer. Sedangkan untuk *potensial customer* menurut (Parvaneh, Abbasimehr, dan Tarokh 2012) terkait pelanggan dengan kategori potensial customer disarankan perusahaan harus menemukan strategi jarak jauh dengan cara menghubungi melalui telepon, email, fax, dan sebagainya untuk memecahkan masalah menggunakan pengaruh seperti iklan informatif.

PRAKATA

Puji syukur kehadiran Allah S.W.T atas segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul “Segmentasi CLV Pelanggan Retail Menggunakan Model LRFM dan Metode AHC Untuk Menentukan Strategi CRM”. Skripsi ini disusun untuk memenuhi salah satu syarat menyelesaikan pendidikan Strata Satu (S1) pada program Studi Informatika Fakultas Ilmu Komputer Universitas Jember.

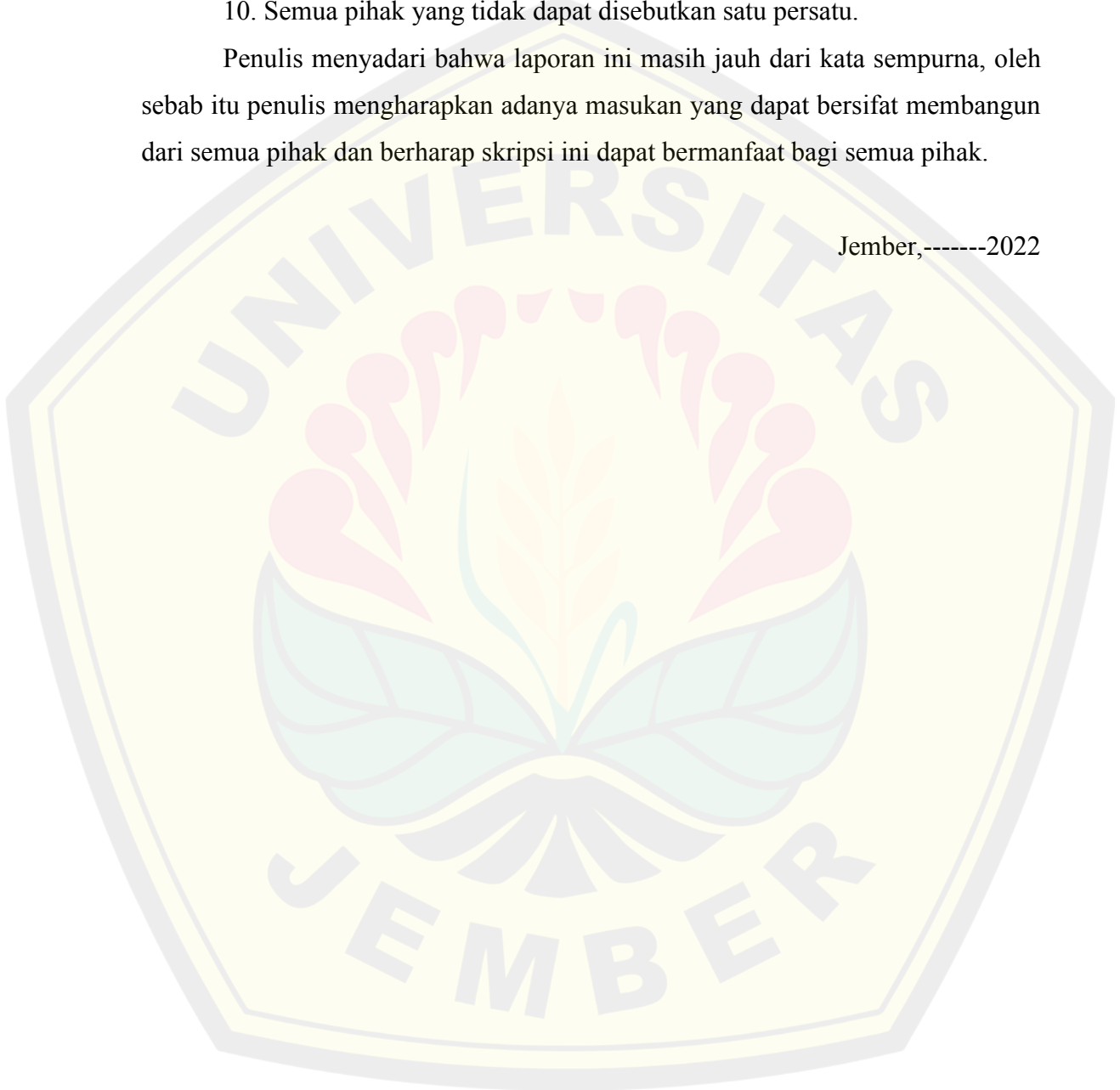
Penyusunan skripsi ini tidak lepas dari dukungan berbagai pihak. Oleh karena itu penulis menyampaikan terima kasih kepada:

1. Drs. Antonius Cahya Prihandoko, M. App.Sc., ph.D., selaku Dekan Fakultas Ilmu Komputer Universitas Jember;
2. Prof. Dr. Saiful Bukhori, ST.,M.Kom selaku Dosen Pembimbing Utama dan Januar Adi Putra,S.Kom.,M.Kom. selaku Dosen Pembimbing Anggota yang telah meluangkan waktu untuk membantu dan mengarahkan serta memberikan saran dalam proses penyelesaian skripsi ini;
3. Prof. Drs.Slamin, M.Comp.Sc.,Ph. selaku Dosen Pembimbing Akademik yang telah mendampingi penulis selama menempuh pendidikan S1;
4. Seluruh Bapak dan Ibu Dosen beserta staf karyawan Fakultas Ilmu Komputer Universitas Jember;
5. Sariyanto, Ibu Endang Lestari yang telah memberikan segalanya dan berdarah untuk putrinya. Pariera Vanesa Purbosari terimakasih telah memberikan kata-kata luar biasa selama mengerjakan skripsi ini, serta keluarga besar yang selalu mendukung dan mendoakan;
6. Yeni, Wulan, Andya, Nur, Fajriah, Mely, Dita, Agnia yang siap menjadi tampungan semua keluh kesah selama mengerjakan skripsi ini;
7. Seluruh orang baik yang saya temui di tahun 2021 hingga sekarang yang memberikan semangat, kekuatan dan menjadi pendengar setia serta memberikan dukungan dalam menyelesaikan skripsi ini;

8. Musisi terhebatku Seventeen yang selalu memberikan kehangatan dan mengingatkan melalui musik agar selalu kuat menghadapi semuanya dan menemani penulis untuk tetap semangat dalam menuntaskan skripsi serta menjadi panutan untuk berjuang dengan keras dengan kekuatan sendiri;
9. Teman-teman grub Keluarga haji bolot dan grub DECISION
10. Semua pihak yang tidak dapat disebutkan satu persatu.

Penulis menyadari bahwa laporan ini masih jauh dari kata sempurna, oleh sebab itu penulis mengharapkan adanya masukan yang dapat bersifat membangun dari semua pihak dan berharap skripsi ini dapat bermanfaat bagi semua pihak.

Jember,-----2022



DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
PERSEMBAHAN.....	ii
MOTTO	iii
PERNYATAAN.....	iv
PENGESAHAN PEMBIMBING.....	v
PENGESAHAN PENGUJI	vi
RINGKASAN	vii
PRAKATA.....	ix
DAFTAR ISI.....	xi
DAFTAR TABEL	xiv
DAFTAR GAMBAR.....	xv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian.....	4
1.5 Batasan Masalah.....	5
BAB 2 TINJAUAN PUSTAKA.....	6
2.1 Penelitian Terdahulu	6
2.2 Retail	8
2.3 Koperasi Pegawai Republik Indonesia UNEJ.....	8
2.4 Customer Relationship Management (CRM)	9
2.4.1 Operational CRM	9
2.4.2 Collaborative CRM	10
2.4.3 Analytical CRM.....	10
2.5 Clustering	10
2.6 Hierarchial <i>Clustering</i>	11
2.7 Agglomerative Hierarchial <i>Clustering</i>.....	11

2.7.1 <i>Single Linkage</i>	12
2.7.2 <i>Complete Linkage</i>	13
2.7.3 <i>Average Linkage</i>	13
2.7.4 <i>Minimum Variance: Ward's Linkage</i>	14
2.8 Sillhoutte Coefficient	15
2.9 Customer Lifetime Value (CLV)	16
2.10 Model LRFM	16
2.11 Normalisasi <i>Min-max</i>	17
BAB 3 METODOLOGI PENELITIAN	18
3.1 Jenis Penelitian	18
3.2 Waktu Penelitian	18
3.3 Tahapan Penelitian	18
3.4 Data Collection	19
3.5 Data Preprocessing	20
3.7 Data Understansing	20
3.8 Pembentukan Model LRFM	20
3.9 Normalisasi Data	21
3.10 Clustering	21
3.11 Pengujian Hasil <i>Clustering</i>	22
3.12 Penentuan Nilai CLV	23
3.13 Analisa Dan Visualisasi Data	23
3.14 Analisa Strategi CRM	27
BAB 4 HASIL DAN PEMBAHASAN	28
4.1 Data Collection	28
4.2 Data Preprocessing	28
4.3 Data Understanding	30
4.3.1 Grafik Penjualan Harian	30
4.3.2 Grafik Jumlah Customer Perbulan	31
4.3.3 Grafik Penjualan Per Jam dan Penjualan Per Hari	32
4.4 Pembentukan Model LRFM	34

4.5 Normalisasi.....	35
4.6 Clustering	36
4.6.1 Clustering Menggunakan Agglomerative Hierarchical Clustering	36
4.6.1 Menentukan <i>cluster</i> dengan melakukan pemotongan dendrogram	37
4.7 Pengujian <i>cluster</i>.....	39
4.8 Visualisasi <i>cluster</i>	41
4.8.1 Visualisasi Data Dengan Menggunakan 3D	41
4.8.2 Visualisasi Data Menggunakan 2D	42
4.9 Perhitungan atau pembobotan <i>clv</i> untuk masing masing <i>cluster</i> ...	43
4.10 Penentuan kelompok atau golongan <i>cluster</i>	44
4.11 Analisa Strategi CRM Terhadap Hasil <i>Cluster</i>	45
4.11.1 Analisa hasil <i>cluster</i> 0.....	45
4.11.2 Analisa hasil <i>cluster</i> 1	47
BAB 5 PENUTUP.....	50
5.1 KESIMPULAN	50
5.2 SARAN	50
DAFTAR PUSTAKA	51
LAMPIRAN.....	55
A. Data Hasil Pembentukan Model LRFM.....	55
B. Visualisasi Data Dengan Menggunakan 3D	55
C. Visualisasi Data Menggunakan 2D	57
D. Source Code.....	59

DAFTAR TABEL

Tabel 3.1 Deskripsi kelompok dan cluster	24
Tabel 4.1 Tabel variable yang dihapus dan tidak digunakan pada penelitian.....	28
Tabel 4.2 Nama Atribut dan Penjelasan atribut yang digunakan	29
Tabel 4.3 Hasil Pemodelan Model LRFM	35
Tabel 4. 4 Hasil Normalisasi Model LRFM Menggunakan MinMax scaler	36
Tabel 4.5 Hasil Perhitungan Silhoutte terhadap kemungkinan jumlah cluster	40
Tabel 4. 6 Hasil perkalian pembobotan dengan nilai ahp pada penelitian terdahulu	44
Tabel 4.7 Tabel Hasil clustering berisikan rata-rata digunakan untuk menentukan symbol LRFM	44
Tabel 4.8 Karakteristik Cluster 0 atau New Customer	466
Tabel 4.9 Hasil Ringkasan analisa cluster 0.....	46
Tabel 4.10 Karakteristik Cluster 1 atau Potensial Customer	468
Tabel 4.11 Hasil Analisa Cluster 1	469

DAFTAR GAMBAR

Gambar 2.1 <i>Single Linkage</i>	13
Gambar 2.2 <i>Complete Linkage</i>	13
Gambar 2. 3 <i>Average Linkage</i>	14
Gambar 2.5 Ward Linkage.....	15
Gambar 3.1 Alur Penelitian.....	19
Gambar 3.2 Alur Agglomerative hierarchical clustering.....	22
Gambar 3.3 Pembobotan LRFM menggunakan AHP oleh Parvaneha, Abbasimehrb, and Tarokhc 2012.....	23
Gambar 3.4 Customer Loyalty Matrix untuk Customer clustering mengacu nilai Length dan Recency.....	24
Gambar 4.1 Grafik Garis Penjualan Harian	31
Gambar 4.2 Grafik Penjualan Perbulan	32
Gambar 4.3 Grafik penjualan per jam dan per hari.....	33
Gambar 4.4 Code Pemodelan Model LRFM	35
Gambar 4.5 Normalisasi Mennggunakan MinMax.....	36
Gambar 4.6 Code Clustering dan Plot Menggunakan Dendogram.....	37
Gambar 4.7 Gambar Hasil Clustering Menggunakan AHC Menggunakan Dendogram.....	38
Gambar 4.8 Gambar pemotongan dendogram untuk 3 cluster	38
Gambar 4.9 Gambar pemotongan dendogram untuk 4 cluster	39
Gambar 4.10 Gambar pemotongan dendogram untuk 6 cluster	39
Gambar 4.11 Code Pengujian Hasil Cluster	40
Gambar 4.12 Gambar Hasil Scatter plot 3D model LRF dimana nilai x = Length_of_stay, y = Recency, z = Frequency.....	41
Gambar 4.13 Gambar Hasil Scatter plot 2D model LR.dimana nilai x = length, y = recency	42
Gambar 4.14 Code Menentukan Nilai CLV untuk masing-masing cluster	43

BAB 1 PENDAHULUAN

1.1 Latar Belakang

Retail merupakan semua kegiatan penjualan barang untuk konsumen akhir baik dalam bentuk eceran, grosir, serta menggunakan pelayanan modern maupun tradisional (Chaniago dan Bandung 2021). Retail modern seperti gerai toko swalayan atau supermarket mengalami peningkatan permintaan (Gunawan 2022). Menurut data dari Bank Indonesia, Penjualan ecer pada bulan April 2022 diperkirakan akan meningkat secara bulanan. Hal ini terlihat dari Indeks Penjualan (IPR) April tahun 2022 yang tercatat sebesar 219,3 atau secara bulanan penjualan tumbuh ke angka 6,8% (Haryono 2022). Berdasarkan data tersebut menunjukkan bahwa toko retail telah mengalami peningkatan penjualan dalam beberapa bulan terakhir.

Peningkatan penjualan perlu dilakukan untuk mencegah terjadinya penurunan penjualan. Pencegahan tersebut dapat dilakukan dengan memperhatikan pelanggan secara teliti yang berguna untuk mengurangi kehilangan pelanggan (Rahmawati, Toana, dan Sofyan 2021). Selain itu, toko retail juga harus mempertahankan pelanggan yang memiliki nilai tinggi terhadap perusahaan (Marisa dkk. 2019). Strategi yang biasa digunakan adalah tidak memperlakukan semua pelanggan dengan cara yang sama. Hal ini dikarenakan memperlakukan semua pelanggan secara setara adalah strategi yang dapat mengubah pelanggan yang tidak terlalu berharga bagi perusahaan menjadi penghancur nilai bagi perusahaan (Monalisa 2018). Maka dari itu diperlukan strategi dimana perusahaan mempertahankan pelanggan dengan menggunakan *Customer Relationsh ip Management* (CRM).

CRM merupakan strategi dimana sebuah perusahaan atau badan usaha yang berfokus pada pelanggan (Widyawati, Saptomo, dan Utami 2020). CRM merupakan strategi bisnis untuk mendapatkan keuntungan, pendapatan, dan kepuasan dari pelanggan serta bagaimana cara memperlakukan pelanggan (Justitia, Hidayat, dan Santoso 2021). Strategi CRM merupakan jembatan penghubung antara perusahaan atau badan usaha dengan pelanggan (Amanah 2017).

Penyusunan strategi penjualan dapat dilakukan dengan melakukan analisa CRM. Analisa CRM ini berkaitan dengan membagi pelanggan dengan sifat dan kriteria yang sama atau disebut juga dengan segmentasi pelanggan (Hananto, Churniawan, dan Wardhanie 2017). Dengan membagi pelanggan perusahaan dapat mengetahui minat belanja masyarakat terhadap toko dan pelanggan yang mempunyai loyalitas belanja di toko (Marisa dkk. 2019). Pelanggan yang memiliki loyalitas kepada perusahaan dapat diketahui dengan menghitung nilai *customer lifetime value* (CLV) dari mengekstrak data transaksi penjualan dan data customer.

CLV adalah nilai yang diberikan perusahaan kepada pelanggan dan nilai yang diberikan pelanggan kepada perusahaan selama periode hubungan (Marisa dkk. 2019). Nilai CLV dapat digunakan sebagai bahan evaluasi nilai masa depan pelanggan kepada perusahaan. Ada banyak model yang dikembangkan untuk mengukur nilai CLV. Beberapa model yang bisa digunakan untuk mengukur CLV adalah *RFM Models*, *Probably Models*, *Econometric Models*, *Persistence Models*, *Computer Science Models*, *Diffusion/Growth Models*. Model RFM dipilih pada penelitian ini karena model RFM sudah banyak digunakan dalam bisnis industri dan retail dan mudah diterapkan pada praktiknya (Gupta dkk. 2006). LRFM merupakan model pengembangan dari RFM oleh Arthur Hughes (1994) dengan penambahan variabel L (*Length*) oleh Chang and Tsay (2004). *Length* merupakan variabel untuk mengukur nilai rentang lamanya transaksi pertama dan terakhir pelanggan. Penambahan variabel *length* (jangka panjang) pada model LRFM digunakan untuk membedakan pelanggan yang memiliki hubungan jangka pendek atau jangka panjang dengan perusahaan (Zoeram dan Mazidi 2018). Penambahan variabel *length* dinilai penting karena dapat mendukung variabel *Recency* pada pelanggan berdasarkan lamanya transaksi terakhir pada periode tertentu (Monalisa 2018). Setelah melakukan pembentukan model LRFM selanjutnya adalah melakukan segmentasi pelanggan dengan menggunakan *machine learning*.

Beberapa metode yang dapat digunakan dalam segmentasi pelanggan adalah K-means, Fuzzy c-means dan *Agglomerative Hierarchical Clustering*. *Agglomerative Hierarchical Clustering* merupakan salah satu teknik tradisional dalam *clustering* analisis (Mongi dkk. 2019). Berbeda dengan metode k-means dan

fuzzy c-means yang menentukan jumlah *cluster* di awal. *Agglomerative Hierarchical Clustering* tidak mengasumsikan jumlah *cluster* terlebih dahulu namun pengelompokkan berdasarkan dengan kesamaan antar jarak atau dissimilarity (Nurseptiani, Satria, dan Burhan 2021). Konsep algoritma ini menggambarkan semua data menjadi satu *cluster* kecil (*leaf*) lalu dikelompokkan dengan *cluster* lain dan membentuk *cluster* besar (*nodes*). Metode *Agglomerative* menghitung nilai (*dis*)similarity dengan menggunakan panjang garis antar data (garis tegak lurus) pada pohon *dendogram* (Müller dan Guido 2017). Sehingga membuat algoritma ini lebih jelas dan persuasif untuk mengelompokkan data (Gu 2021).

Berdasarkan penjabaran yang telah dipaparkan. Metode yang digunakan pada penelitian ini adalah *Agglomerative Hierarchical Clustering*. Metode ini dipilih untuk segmentasi *customer lifetime value* pelanggan di toko retail dengan menggunakan model LRFM untuk menentukan strategi promosi kepada pelanggan berdasarkan CRM. Penggunaan metode *Agglomerative Hierarchical Clustering* ini terbukti pada penelitian Rahyana Putri Justitia, Nurul Hidayat, Edy Santoso pada tahun 2021 dengan judul “Implementasi Metode *Agglomerative Hierarchical Clustering* pada segmentasi pelanggan barbershop”. Segmentasi pelanggan menggunakan metode *Agglomerative Hierarchical Clustering* memberikan nilai akurasi berdasarkan parameter jarak terbaik dengan rata – rata nilai paling tinggi yaitu 0.968850698 (Justitia, et.,al,2021). Selain itu pada penelitian lain milik Crisnandra Rahmita Mardiantien, Imelda Atastina, Ibnu Abror mengenai *Product Segmentation Based on Sales Transaction Data Using Agglomerative Hierarchical Clustering* memberikan jumlah *cluster* sebanyak 7 dengan perhitungan jarak terbaik dengan menggunakan *Ward’s Linkage*. Dengan penerapan segmentasi *customer lifetime value* di toko retail ini diharapkan dapat membantu toko retail untuk menentukan strategi promosi yang sesuai dengan kelompok pelanggan yang sama.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka hasil rumusan masalah yang terbentuk adalah sebagai berikut:

1. Bagaimana hasil segmentasi *customer lifetime value* pelanggan retail berdasarkan model LRFM dengan menggunakan metode *Agglomerative Hierarchical Clustering*?
2. Bagaimana strategi CRM yang dapat diberikan pada toko retail berdasarkan hasil segmentasi?

1.3 Tujuan Penelitian

Tujuan dalam laporan ini antara lain:

1. Mengsegmentasi *customer lifetime value* pelanggan retail berdasarkan model LRFM dengan menggunakan metode *Agglomerative Hierarchical Clustering*.
2. Menentukan strategi CRM yang dapat diberikan pada toko retail berdasarkan hasil segmentasi.

1.4 Manfaat Penelitian

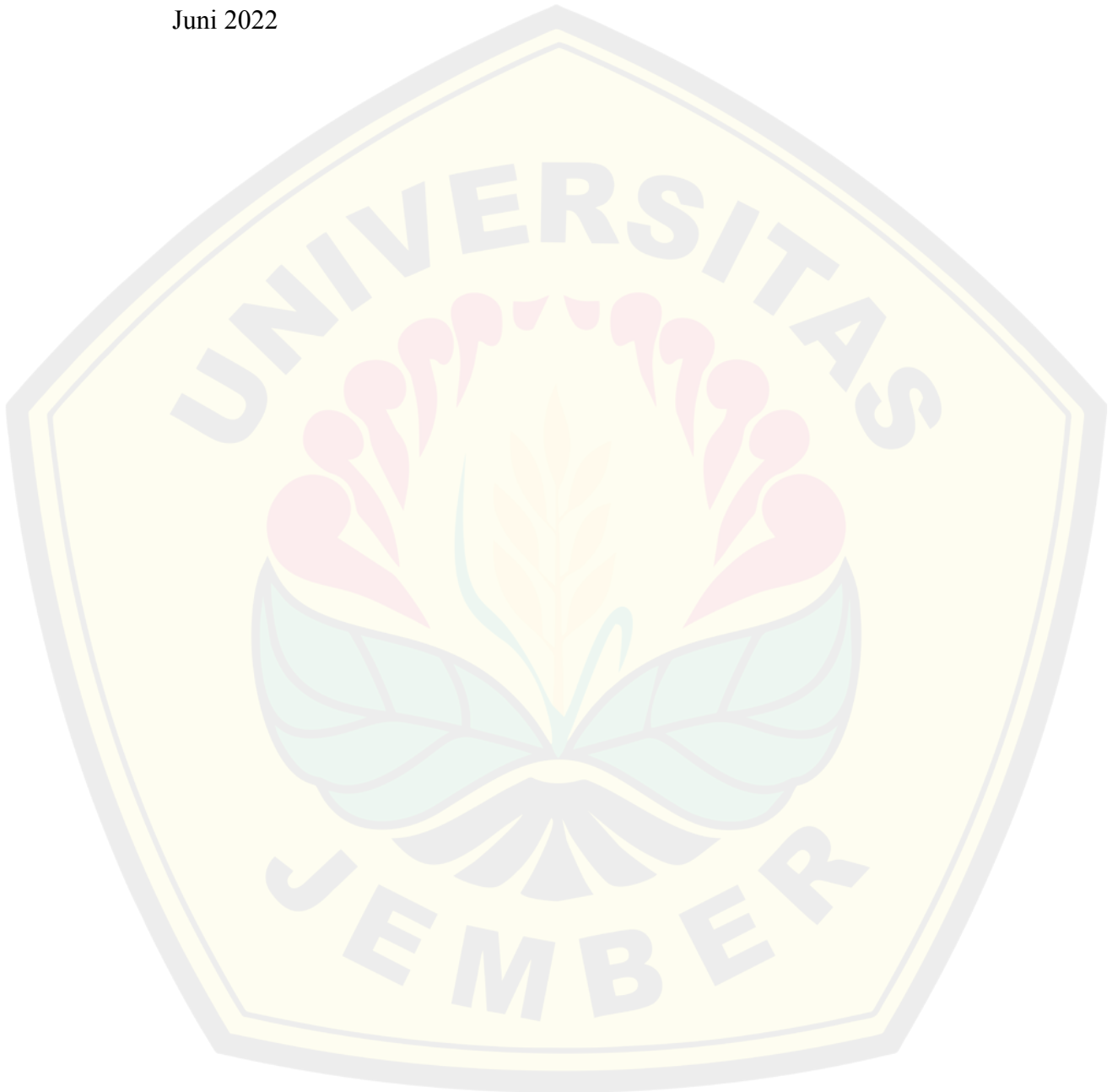
Manfaat dari penelitian ini adalah sebagai berikut:

1. Bagi Akademik
Hasil dari penelitian diharapkan dapat menambah wawasan pengetahuan. Penelitian ini juga diharapkan dapat digunakan sebagai referensi terkait dengan penelitian selanjutnya dan dapat memberikan kontribusi yang nyata terhadap teori-teori yang berkaitan dengan permasalahan yang diangkat oleh peneliti.
2. Bagi Objek Penelitian
Hasil dari penelitian dapat digunakan sebagai bahan kajian untuk menentukan strategi penjualan dan evaluasi dari sistem operasi minimarket. sehingga dapat membantu pihak objek penelitian untuk memperbaiki kendala yang terjadi object penelitian.
3. Bagi Peneliti
Dapat menambah ilmu, wawasan baru yang lebih luas mengenai teori dan penerapan *clustering*, *machine learning*, pengolahan data dan analisa perilaku pelanggan.

1.5 Batasan Masalah

Agar penelitian lebih terfokuskan untuk mencapai sasaran dan tujuan yang diharapkan, maka permasalahan yang ada hanya dibatasi pada:

1. Data yang digunakan adalah data transaksi penjualan KPRI UNEJ
2. Data transaksi selama 6 bulan terhitung dari bulan januari tahun 2022 hingga Juni 2022



BAB 2 TINJAUAN PUSTAKA

Bagian ini memaparkan beberapa pustaka yang berkaitan dengan masalah yang dibahas, teori terkait masalah dan juga penelitian terdahulu yang digunakan sebagai acuan atau landasan teori dalam melakukan penelitian agar penelitian tidak melenceng dari kaidah ilmu yang ada.

2.1 Penelitian Terdahulu

Penelitian pertama “Klatisasi *Customer Lifetime Value* Dengan Model LRFM Menggunakan Model Algoritma K-Means” oleh Siti Monalisa tahun 2018 (Monalisa 2018). Pada penelitian ini dikatakan bahwa masing-masing pelanggan memiliki kebutuhan, preferensi, harapan, dan perilaku yang berbeda kepada perusahaan. Perusahaan sebaiknya tidak memperlakukan seluruh pelanggan dengan cara yang sama. Hal ini dikarenakan perlakuan tersebut dapat mengakibatkan pelanggan yang tidak begitu bernilai menjadi penghancur nilai perusahaan. Maka pada penelitian ini dilakukan segmentasi CLV pelanggan untuk mengetahui kelompok yang memiliki nilai pelanggan terhadap perusahaan. Salah satu model yang digunakan untuk mengukur CLV pada penelitian ini adalah LRFM. Metode segmentasi yang digunakan dalam penelitian ini adalah k-means dengan penentuan *cluster* terbaik menggunakan *Dunn Index* dan *Silhouette Coefficient*. Data yang digunakan pada penelitian ini adalah data pelanggan HPAI cabang Pekanbaru dengan data sebanyak 1.867 data pelanggan. Penelitian ini menghasilkan 3 *cluster* yang terbaik dengan hasil kluster 2 nilai CLV 0,208, kluster 3 nilai CLV 0,082 dan kluster 1 dengan nilai 0,038.

Penelitian kedua dengan judul “*Segmentation Model of Customer Lifetime Value in Small and Medium Enterprise (SMEs) using K-Means Clustering and LRFM Model*” oleh Fitri Marisa 2019. Penelitian ini hampir mirip dengan sebelumnya mengenai segmentasi pelanggan. Namun penentuan *cluster* terbaik pada penelitian ini menggunakan metode *elbow*. Data yang digunakan dalam penelitian ini adalah dataset penjualan UKM atau SMEs pulsa semua operator selama periode 1 Januari sampai 30 Juni 2018. Hasil segmentasi menggunakan

Elbow dan *Sum Square Error* (SSE) adalah 2 *cluster* terbaik. Kluster 2 dengan nilai 0.362 dan kluster 1 dengan nilai 0.122.

Adapun penelitian selanjutnya yaitu “Product Segmentation Based on Sales Transaction Data Using Agglomerative Hierarchical *Clustering* and FMC Model (Case Study : XYZ Company) oleh (Mardiantien, Atastina, dan Asror 2020). Perusahaan XYZ merupakan perusahaan yang bergerak dibidang obat-obatan, data transaksi pada perusahaan ini terbilang besar dan terjadi penumpukan. Pada penelitian ini menggunakan model FMC untuk mengetahui insights produk mana yang memiliki kecenderungan dibeli oleh customer. Algoritma Agglomerative Hierarchical *Clustering* ini dipilih karena algoritma ini disebut cukup populer dan tahan terhadap noise data. Dari penelitian ini dihasilkan bahwa AHC dengan metode pengujian CHI dengan hasil kluster terbaik menggunakan jarak hitung Ward’s Linkage. Data pada penelitian ini terbilang cukup besar dimana data transaksi pada Januari-March 2019 terdapat 653.718 data dengan 116 atribut, dan Januari-March 2018 dengan 528.472 data dengan 116 atribut. Hasil penelitian dengan menggunakan Agglomerative Hierarchical *Clustering* ini, menghasilkan delapan *cluster* dengan lima kombinasi FMC untuk data tahun 2018 dan empat *cluster* pada tahun 2019.

Penelitian selanjutnya berjudul “Penerapan *Agglomerative Hierarchical Clustering* untuk segmentasi pelanggan” oleh Widyawati tahun 2020. Permasalahan pada penelitian ini adalah toko komputer yang belum memiliki strategi khusus dan strategi yang diatur belum tepat sasaran. Penelitian ini menggunakan model RFM karena perusahaan ingin mengetahui nilai profitabilitas pelanggan terhadap perusahaan dari transaksi yang telah dilakukan. Penelitian ini menghasilkan 7 *cluster* dengan *cluster* 1, 20 anggota, *cluster* 2, 43 anggota, *cluster* 3, 75 anggota, *cluster* 4, 158 anggota, *cluster* 5, 9 anggota, *cluster* 6, 2 anggota dan *cluster* 7, 1 anggota. Penelitian ini menyarankan untuk penelitian selanjutnya menggunakan metode lain selain *Average Linkage*. Misalnya, penelitian selanjutnya bisa menggunakan *Complete Linkage*, *Single Linkage*, maupun *Ward’s Linkage*. Metode pengukuran jarak juga dapat menggunakan *Euclidean distance* ataupun yang lain.

Adapun jurnal terakhir yang dijadikan sebagai acuan adalah “Implementasi Metode *Agglomerative Hierarchical Clustering* Pada Segmentasi Pelanggan Barbershop (Studi Kasus: Richdjoe Barbershop Malang)” oleh Rhayhana Putri Justitia tahun 2021. Metode pada penelitian dipilih karena *Agglomerative Hierarchical Clustering* merupakan algoritma yang menggabungkan data dengan kemiripan data. Data yang diolah pada penelitian ini adalah data pelanggan periode bulan mei-oktober 2020. Variabel data pada penelitian ini adalah nomor ID, jumlah kedatangan pelanggan berdasarkan nomor ID pada kurun waktu tertentu, jumlah uang yang dikeluarkan setiap pelanggan, dan akumulasi jarak transaksi dari terlama hingga terkini berdasarkan nomor ID. Hasil dari penelitian ini adalah mengetahui keakuratan metode *Agglomerative Hierarchical Clustering* dengan uji validasi menggunakan *Silhouette Coefficient*. Nilai rata-rata jarak tertinggi yaitu pada *Single Linkage* dan *Average Linkage* dengan nilai 0.968850698. saran pada penelitian ini untuk penelitian selanjutnya bisa menggunakan parameter ukur jarak atau metode lain yang berbeda untuk meningkatkan akurasi yang lebih tinggi. Selain itu penelitian selanjutnya juga dapat menggunakan lebih banyak fitur dan jumlah data yang lebih besar, untuk meningkatkan kompleksitas data yang digunakan selama segmentasi

2.2 Retail

Bisnis retail adalah sebuah usaha yang dijalankan oleh sekelompok orang atau seseorang secara individu, menjual berbagai kebutuhan harian masyarakat dan dijalankan secara modern atau tradisional (Chaniago dan Bandung 2021). Retail adalah usaha bisnis yang menjual barang dalam jumlah kecil untuk konsumen akhir. Kegiatan retail bisa dilakukan oleh satu orang atau sekelompok orang, baik secara tradisional atau modern (Chaniago dan Bandung 2021). Berikut merupakan contoh bisnis retail:

2.3 Koperasi Pegawai Republik Indonesia UNEJ

Koperasi merupakan badan usaha yang beranggotakan orang-orang atau badan hukum dengan melandaskan kegiatannya berdasarkan prinsip koperasi

sekaligus sebagai gerakan ekonomi rakyat yang berdasarkan atas asas kekeluargaan (Soejono, I, 2001). Koperasi Pegawai Republik Indonesia UNEJ (KPRI UNEJ) didirikan pada tanggal 2 Agustus 1979 yang berada di jalan kalimantan no. 27 Kabupaten Jember Propinsi Jawa Timur, dengan nomor badan hukum 4388BHII '80. Nomor badan hukum ini diperoleh pada tanggal 12 Februari 1980. KPRI UNEJ pertama kali didirikan dengan nama Koperasi Pegawai Negeri Universitas Jember (KPN UNEJ). KPRI UNEJ membagi usahanya dalam 3 unit usaha yaitu, Unit usaha jasa, Unit usaha simpan pinjam, Unit usaha toko (Khotimah 2015).

2.4 *Customer Relationship Management (CRM)*

CRM adalah proses strategi bisnis untuk memilih pelanggan yang paling menguntungkan perusahaan melayani dan membentuk interaksi antara perusahaan dan pelanggan ini (Reinartz, V, dan J 2006). Tujuannya adalah untuk mengoptimalkan saat ini dan masa depan pelanggan bagi perusahaan (Reinartz, V, dan J 2006). Selain itu CRM memberikan peluang yang lebih baik bagi para pelaku industri bisnis untuk menggunakan informasi yang tersedia untuk mengetahui jenis pelanggan dan menciptakan nilai tambah bagi setiap pelanggan (Rahadian dan Syairudin 2020).

Secara umum, CRM merupakan cara untuk memaksimalkan keuntungan bisnis dengan meningkatkan kepuasan pelanggan dan mempertahankan hubungan dengan pelanggan. Secara garis besar, CRM dikelompokkan dalam 3 jenis (Suryanto 2013) yaitu :

2.4.1 *Operational CRM*

CRM *operational* mengambil alih tugas-tugas administrasi yang berkaitan dengan pemasaran dan penjualan yang selama ini dikerjakan oleh manusia. Tugas-tugas yang diambil oleh sistem ini mencakup optimalisasi dari keseluruhan proses bisnis, baik yang berinteraksi secara langsung ataupun tidak secara langsung dengan pelanggan.

2.4.2 *Collaborative CRM*

Collaborative CRM merupakan jenis CRM yang memecah data silo sesuai dengan kebutuhan perusahaan dan tiap-tiap departemen dalam satu sistem. Contoh aplikasi yang digunakan sebagai CRM *collaborative* adalah e-mail, *personalized publishing*, *e-communities*, forum diskusi, dan sarana lainnya yang dirancang untuk memfasilitasi interaksi antara pelanggan dan pihak perusahaan.

2.4.3 *Analytical CRM*

Analytical CRM merupakan jenis CRM yang memungkinkan memperoleh wawasan tentang *customer*. *Analytical CRM* mengumpulkan informasi pelanggan dari berbagai sumber data untuk menemukan pola yang tidak terlihat. Cara ini membantu perusahaan dalam mengidentifikasi, memahami, memanfaatkan tren dan perilaku pelanggan. Setelah mengetahui pola tersebut perusahaan dapat menyusun kampanye pemasaran serta meningkatkan layanan pelanggan.

Dalam menerapkan strategi CRM dibutuhkan tahapan yang akan dilakukan, menurut Strauss (2001:297) menentukan strategi CRM terdiri dari 3 tahap yaitu mengidentifikasi *customer*, membedakan *customer* dan menyesuaikan *marketing*. Mengidentifikasi *customer* tahapan ini merupakan proses untuk mengetahui mengenai individual informasi *customer*. Tahap kedua membedakan *customer* dimana *customer* yang memiliki informasi yang berada dikelompokkan kedalam kelompok yang sama, misalnya seperti mengidentifikasi *high-value customer* dan mengelompokkan *customer* dengan *value*-nya masing-masing. Mengelompokkan *customer* ini juga merupakan proses segmentasi pelanggan. Tahap ketiga merupakan menyesuaikan *marketing*. Setelah membedakan *customer* berdasarkan informasi yang ada perusahaan menyesuaikan penawaran mereka terhadap kelompok atau individu yang berbeda.

2.5 *Clustering*

Clustering adalah membagi titik atau data kedalam kelompok-kelompok alami berdasarkan kemiripan data atau titik dalam suatu object (Mardiantien,

Atastina, dan Asror 2020). Karakteristik *cluster* tergantung data, ada berbagai jenis *clustering* seperti *representative-based*, *hierarchical*, *density-based*, *graph-based*, dan *spectral clustering* (Zaki dan Meira 2020). *Clustering* termasuk dalam *unsupervised learning* tanpa pengawasan karena tidak memerlukan pembelajaran atau pelatihan dataset untuk mempelajari parameter model (Zaki dan Meira 2020). Metode yang digunakan dalam melakukan segmentasi pelanggan adalah metode *clustering*. Menurut Han, Kamber tahun 2012, *Clustering* dapat disebut juga sebagai segmentasi data pada beberapa aplikasi karena mengelompokkan data besar ke dalam kelompok yang memiliki kemiripan yang sama (Monalisa 2018). Segmentasi data atau pelanggan menggunakan teknik *clustering* untuk mengevaluasi pembuatan strategi bisnis (Mardiantien, Atastina, dan Asror 2020).

2.6 ***Hierarchical Clustering***

Hierarchical clustering adalah teknik *clustering* yang berbentuk hirarki dengan berdasarkan tingkatan tertentu sehingga *clustering* ini menyerupai struktur pohon (Ramadhani, Purnamasari, dan Amijaya 2018). Metode pengelompokan hirarkhis biasanya digunakan apabila nilai *cluster* belum ditentukan atau belum ada informasi jumlah *cluster* yang akan dibuat (Wijuniamurti, Nugroho, dan Rachmawati 2022). Ada 2 pendekatan algoritma utama yang digunakan dalam *Hierarchical Clustering* yaitu *Agglomerative* dan *disive* (Wijuniamurti, Nugroho, dan Rachmawati 2022). *Agglomerative* adalah pengelompokan hirarki yang bekerja dari bawah ke atas. *Divisive* adalah pengelompokan hirarki yang bekerja dari atas ke bawah (Zaki dan Meira 2020). *Agglomerative Hierarchical clustering* menggabungkan setiap data kecil hingga membentuk satu kelompok besar. Sedangkan *hierarchical divisive clustering* memisahkan data yang terbentuk dalam kelompok besar menjadi setiap data kecil (Bachtiar, Dharmayanti, dan Hamzah 2017)

2.7 ***Agglomerative Hierarchial Clustering***

Agglomerative Hierarchical Clustering adalah salah satu teknik pengelompokan data atau *clustering*. Teknik ini akan mengelompokkan data

secara berulang-ulang dari data diasumsikan sebagai *cluster* kecil ke *cluster* besar berdasarkan kemiripan data satu dengan data lain dan membentuk sebuah hirarki (Justitia, Hidayat, dan Santoso 2021). Langkah utama dalam menentukan pasangan *cluster* adalah dengan menentukan pasangan terdekat dengan berdasarkan nilai matriks jarak. Pasangan terdekat ini bisa ditentukan dengan melihat rata-rata nilai jarak pada setiap observasi (Widyawati, Sptommo, dan Utami 2020). Jarak antar-*cluster* pada metode *Agglomerative Hierarchical Clustering* didasarkan pada jarak antara dua titik, yang biasanya menggunakan jarak *euclidean distance* dengan persamaan sebagai berikut:

$$\| \mathbf{x} - \mathbf{y} \| = \left(\sum_{i=1}^d (x_i - y_i)^2 \right)^{\frac{1}{2}} \quad (1)$$

Keterangan:

$\| \mathbf{x} - \mathbf{y} \|$ = Jarak *Euclidean* antara vektor \mathbf{x} dan \mathbf{y}

d = banyak peubah/ parameter yang digunakan

x_i = Object \mathbf{x} data ke i

y_i = Object \mathbf{y} data ke i

Berikut beberapa metode penentuan jarak dalam teknik pengelompokan *Agglomerative Hierarchical Clustering* yang digunakan untuk mencari pasangan terdekat menurut (Zaki dan Meira 2020) :

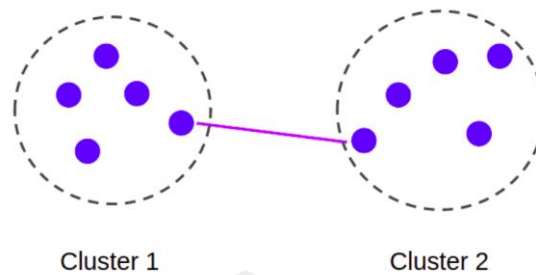
2.7.1 *Single Linkage*

Single Linkage atau *Nearest Neighbor* merupakan sebuah skala jarak berdasarkan jarak paling kecil tiap data. Algoritma jarak *Single Linkage* dimulai dengan memilih jarak terkecil atau terdekat dalam suatu matriks data. Penggambaran bentuk *cluster* dapat dilihat pada gambar 1.

$$\delta(C_i, C_j) = \min\{\| \mathbf{x} - \mathbf{y} \| \mid \mathbf{x} \in C_i, \mathbf{y} \in C_j\} \quad (2)$$

Keterangan:

$\delta(C_i, C_j)$ = Jarak *Single Linkage* antara *cluster* i dan *cluster* j

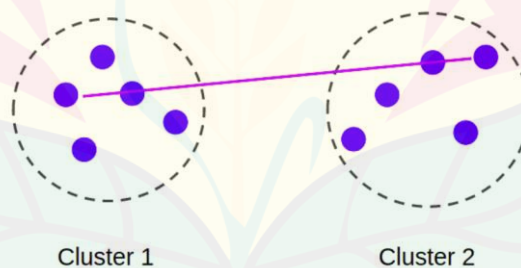
Gambar 2.1 *Single Linkage*2.7.2 *Complete Linkage*

Complete Linkage atau *Furthest Neighbor* merupakan sebuah skala jarak berdasarkan jarak paling besar atau paling jauh tiap data. Skala jarak pada algoritma ini dimulai memilih jarak terbesar atau paling jauh dalam matriks data. Penggambaran bentuk *cluster* dapat dilihat pada gambar 2.

$$\delta(C_i, C_j) = \max\{\| \mathbf{x} - \mathbf{y} \| \mid \mathbf{x} \in C_i, \mathbf{y} \in C_j\} \quad (3)$$

Keterangan:

$\delta(C_i, C_j)$ = Jarak *Complete Linkage* antara *cluster i* dan *cluster j*

Gambar 2.2 *Complete Linkage*2.7.3 *Average Linkage*

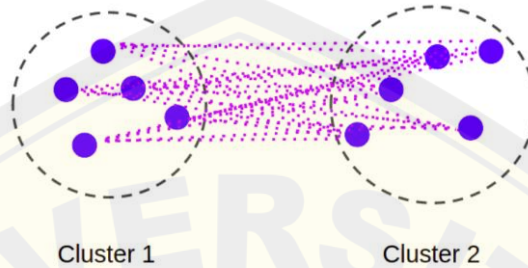
Group average atau *Unweighted Pair Group Method with Arithmetic mean* merupakan jarak antara dua *cluster* dengan definisi sebagai jarak rata-rata pasangan antara dua *cluster*. Penggambaran bentuk *cluster* dapat dilihat pada gambar 3.

$$\delta(C_i, C_j) = \frac{\sum_{x \in C_i} \sum_{y \in C_j} \|x - y\|}{n_i, n_j} \quad (4)$$

Keterangan:

$\delta(C_i, C_j)$ = Jarak *Average Linkage* antara *cluster i* dan *cluster j*

n_i, n_j = Semua titik dalam *cluster C_i* atau jumlah nilai data dari setiap *cluster*



Gambar 2.3 *Average Linkage*

2.7.4 *Minimum Variance: Ward's Linkage*

Ward's linkage atau *Minimal Increase of Sum-of-Squares* merupakan jarak antara dua *cluster* dengan menghitung jumlah kesalahan kuadrat atau *Sum of Squared Errors* (SSE). Dalam metode ini jarak antara dua *cluster* adalah jumlah kuadrat antara dua *cluster* untuk seluruh variabel. Metode Ward menggabungkan antara *cluster* yang meminimalkan peningkatan SSE. Dimana digambarkan dalam rumus dengan nilai n_1 dan n_2 didapat dari $\frac{2}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2n_1n_2}{n_1+n_2}$.

$$\begin{aligned} \delta(C_i, C_j) &= \Delta SSE_{ij} \\ &= \left(\frac{n_i n_j}{n_i + n_j} \right) \| \mu_i - \mu_j \|^2 \end{aligned} \quad (5)$$

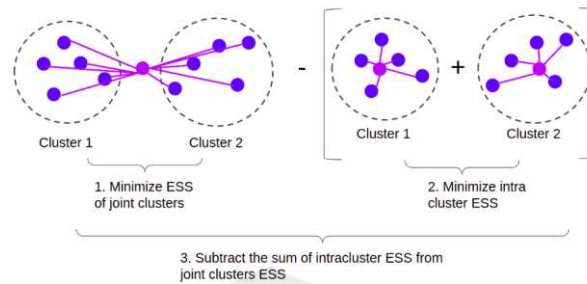
Keterangan:

$\delta(C_i, C_j)$ = Jarak *Ward's linkage* antara *cluster i* dan *cluster j*

ΔSSE_{ij} = Besaran perubahan nilai *Sum of Square Error* dari i, j

n_1, n_2 = Nilai dari $\frac{2}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2n_1n_2}{n_1+n_2}$

μ_i, μ_j = Nilai dari $\frac{1}{n_i} \sum_{x \in C_i} x$



Gambar 2.4 Ward Linkage

2.8 *Silhouette Coefficient*

Silhouette Coefficient merupakan salah satu method dalam melakukan evaluasi yang berguna untuk mengetahui kualitas dan kekuatan suatu *cluster* beserta suatu data dalam *cluster* tersebut (Handoyono, Mangkudjaja, Nasution, 2014). Pengujian silhouette coefficient dilakukan untuk melihat kualitas hasil pengelompokan masing-masing perhitungan jarak, maka perlu dilakukan uji homogenitas.(Hidayati dkk. 2021). Pengujian dilakukan setelah mencapai konvergensi 0 di mana hasil pengelompokan terakhir sama dengan pengelompokan sebelumnya. Dengan kata lain, tidak ada data yang berpindah cluster. Silhouette sendiri merupakan gabungan dari dua metode yaitu metode *Cohesion* berfungsi mengukur seberapa dekat relasi antar elemen atau object dalam suatu *cluster* dan metode separation berfungsi untuk mengukur seberapa jauh relasi anatar elemen atau object dalam suatu *cluster* (Pramesti, Furqon, dan Dewi 2017). Rumus Silhouette Coefficient dapat dilihat pada persamaan 6

$$SC = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (6)$$

Keterangan:

SC = nilai silhouette semua *cluster*

a_i = Jarak rata rata antara i dan semua titik data lain dalam *cluster* i berada

b_i = jarak rata rata minimum dari i semua *cluster* dimana i tidak termasuk

Nilai hasil silhouette coefficient berada pada range nilai -1 hingga 1. Semakin nilai silhouette coefficient mendekati 1, maka semakin baik

pengelompokkan data dalam satu *cluster*. Jika nilai silhouette coefficient mendekati -1, maka semakin buruk pengelompokkan data didalam satu *cluster*.

2.9 Customer Lifetime Value (CLV)

Customer Lifetime Value (CLV) adalah metode yang menghubungkan nilai yang diberikan perusahaan kepada pelanggan atau nilai pelanggan kepada perusahaan dalam masa periode hubungan (Marisa dkk. 2019). Menurut Kumar dkk (2006) CLV juga dapat diartikan sebagai nilai sekarang dari keuntungan masa depan yang dihasilkan dari pelanggan selama masa bisnis dengan perusahaan. Perhitungan nilai CLV pelanggan pada setiap segmen pelanggan dapat digunakan untuk membantu perencanaan pemasaran kepada pelanggan, penetapan perencanaan pemasaran berdasarkan segmentasi pelanggan dinilai menjadi lebih efektif serta mempertahankan pelanggan. pembangunan relasi yang baik dengan pelanggan akan membuat *re-design* bisnis dan *loyalty* pelanggan sehingga akan berpengaruh pada peningkatan profitabilitas perusahaan (Lohonauman 2020).

2.10 Model LRFM

Model RFM merupakan model yang biasanya digunakan untuk menentukan segmentasi pelanggan. RFM digunakan untuk merepresentasikan karakteristik kebiasaan pelanggan metode ini kembangkan oleh Arthur Hughes (1994) (Monalisa 2018). Model ini mengalami perpanjangan variabel yaitu *length* oleh Chang dan Tsay (2004). Penambahan *variable length* digunakan untuk membedakan hubungan jangka panjang pelanggan dengan perusahaan. Model LRFM melibatkan 4 variabel yaitu *Length of the purchases*, *Recency of the last purchases*, *Frequency of the purchases*, dan *Monetary value of the purchases* (Marisa dkk. 2019).

Length merupakan rentang waktu (jarak waktu) hubungan antara perusahaan dan pelanggan. nilai ini diambil dari awal transaksi pembelian dan terakhir dari pelanggan terhadap toko dalam skala waktu penelitian. *Recency* adalah rentang waktu terakhir pelanggan dalam melakukan transaksi pembelian pada akhir skala waktu penelitian. *Frequency* didefinisikan sebagai seberapa sering pelanggan

melakukan transaksi pembelian selama skala waktu penelitian. *Monetary* adalah jumlah nominal transaksi penjualan untuk setiap pelanggan dalam yang melakukan pembelian selama waktu penelitian.

2.11 Normalisasi *Min-max*

Normalisasi data atau data *transformation*, normalisasi dilakukan agar tidak terjadi perbedaan nilai yang begitu tinggi antar variabel sehingga nilai pada sebaran atribut adat menjadi seimbang. Selain itu, Normalisasi sebelum *clustering* umumnya penting dilakukan pada metode yang menggunakan jarak, seperti jarak Euclidean (Mohamad dan Usman 2013). Tujuan lain dari normalisasi adalah untuk mencegah fitur yang lebih besar daripada dengan jumlah yang lebih kecil. Hal ini dilakukan untuk menyamakan dimensi atau besaran variabilitas sebuah fitur. Sehingga perbedaan nilai model LRFM pada *atribut Monetary* pada variabel *length*, *Recency*, dan *Frequency* tidak jauh (Pratikto dan Damastuti 2021). Salah satu teknik normalisasi yang dapat digunakan adalah normalisasi *min-max*. Normalisasi Min-max merupakan proses pengambilan data dimana data yang diambil akan diukur dalam unit tekniknya, dengan mengubah data menjadi nilai dengan skala antara 0,0 hingga 1,0. Secara matematis perhitungan min-max dapat menggunakan rumus sebagai berikut:

$$X_{norm} = \frac{x' - \min(x)}{\max(x) - \min(x)} (new_{\max}(x) - new_{\min}(x)) + new_{\min}(x) \quad (6)$$

Dimana:

x	= Atribut data
$\min(x)$ dan $\max(x)$	= Nilai absolut minimal dan maksimal dari x
x'	= Nilai lama dari setiap entri dalam data
$(new_{\max}(x))$	= Nilai Maximum dari range
$new_{\min}(x)$	= Nilai Minimum pada skala baru

BAB 3 METODOLOGI PENELITIAN

Pada bagian metode penelitian ini menjelaskan mengenai metode penelitian yang digunakan. Metode penelitian ini digunakan untuk menggambarkan tahapan penelitian, memahami proses bisnis di toko retail terkait, mengolah data, menganalisa data serta memberikan saran terkait berupa strategi promosi.

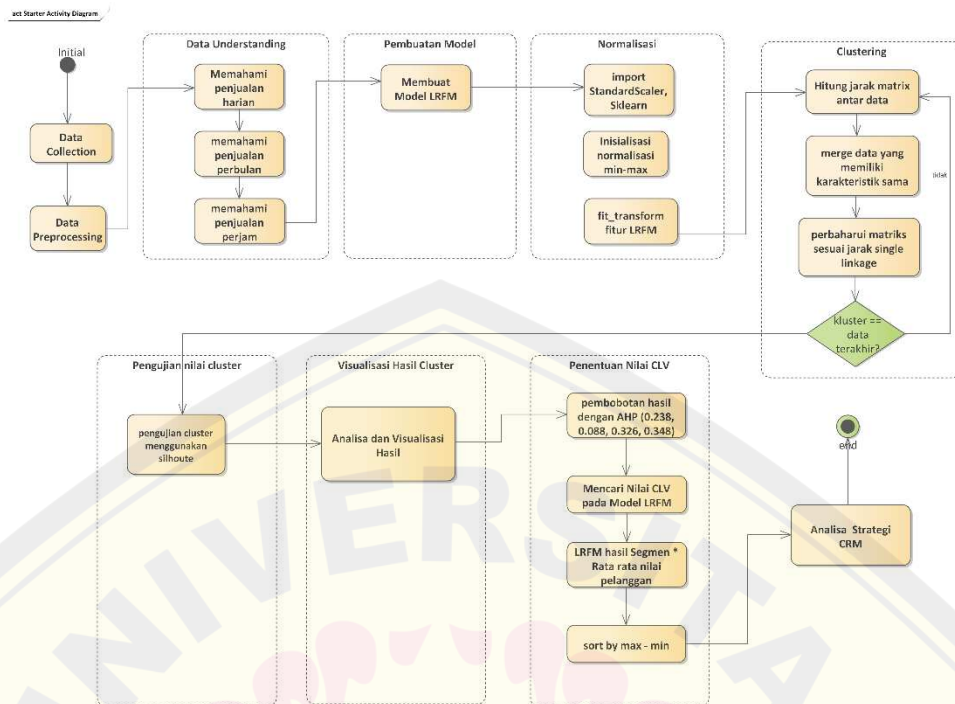
3.1 Jenis Penelitian

Jenis penelitian yang dilakukan merupakan jenis penelitian terapan atau (*applied research*). Penelitian terapan adalah penelitian yang dilakukan secara khusus dan digunakan untuk memecahkan masalah serta membantu menentukan suatu keputusan atau kebijakan baru (Noor, 2011). Tujuan penelitian terapan adalah pemecahan masalah dengan hasil penelitian dapat di aplikasikan untuk kepentingan manusia, baik secara individu, kelompok, ataupun keperluan industri tidak hanya terbatas pada wawasan keilmuan saja (Iriana 2017).

3.2 Waktu Penelitian

Penelitian ini dilakukan di Toko KPRI UNEJ Jl. Sumatra No.105, Tegal Boto Lor, Sumbersari, Kec. Sumbersari, Kabupaten Jember. Pengambilan data daTahapan Penelitian

Alur tahapan penelitian digambarkan pada diagram gambar 3.3.1 sebagai berikut;



Gambar 3.1 Alur Penelitian

3.3 Data Collection

Tahapan pengumpulan data bertujuan untuk mendapatkan data recap yang dibutuhkan dalam penelitian. Data collection pada penelitian ini adalah dengan meminta data transaksi penjualan dari pihak KPRI UNEJ. Data transaksi penjualan dari KPRI UNEJ ini dalam bentuk bulan mulai dari bulan November 2021 hingga April 2022. Berdasarkan data digunakan dalam penelitian terdapat 2 sumber data yaitu data sekunder dan data primer. Data sekunder merupakan data atau informasi tambahan yang tidak didapat langsung dari perusahaan. Data sekunder ini didapatkan dari jurnal-jurnal, skripsi terkait dengan penelitian. Sedangkan data primer ialah data hasil wawancara dan penelitian secara langsung kepada pihak perusahaan. Data primer ini merupakan data transaksi penjualan, data customer, dan wawancara lanjut terkait alur bisnis atau toko retail, serta permasalahan recap didalamnya.

3.4 Data Preprocessing

Data *preprocessing* ini merupakan tahapan dimana data diubah menjadi bentuk data *frame* (*Tabel*). Hal ini dilakukan untuk mengurangi kemungkinan data berbentuk array. Selanjutnya data diolah berdasarkan keperluan penelitian dengan menghapus variabel yang tidak digunakan. Penghapusan variabel yang tidak digunakan seperti pajak dan keterangan. Dan mempertahankan data variabel yang akan digunakan seperti no transaksi, tanggal, kode barang, jumlah, sub total, kode anggota, alamat anggota, penjualan. Tahapan preprocessing ini juga menghasilkan LRFM sebagai model data yang digunakan dalam proses segmentasi.

3.7 Data Understanding

Data understanding merupakan tahapan dimana melakukan pemahaman pada data sebelum diolah lebih lanjut. Penelitian ini akan dilakukan pemahaman data seperti bagaimana penjualan toko selama November 2021 hingga April 2022. Hal lain yang dapat dilakukan pada tahap data *understanding* adalah membiasakan diri terhadap variabel-variabel data pada data penelitian. Mengidentifikasi masalah, kualitas data untuk menemukan wawasan baru mengenai sebuah data atau juga untuk mendeteksi subset atau variabel menarik yang dapat digunakan untuk membentuk informasi baru terhadap sebuah data tersebut. Selain itu tahapan data understanding juga dilakukan untuk mengetahui variabel apa yang dapat digunakan untuk membentuk model LRFM.

3.8 Pembentukan Model LRFM

Model LRFM merupakan perkembangan dari model RFM dengan penambahan variabel L atau *length*. Model LRFM adalah model yang akan digunakan untuk menentukan segmentasi pelanggan dan merepresentasikan karakteristik kebiasaan pelanggan dalam sebuah transaksi pembelian di sebuah perusahaan. Model LRFM ini dibentuk dari variabel yang telah dipahami pada tahap data understanding. Variabel yang digunakan untuk membentuk model LRFM

adalah kode pelanggan, nomer/struk pembelian barang, total penjualan atau total uang yang dilakukan dalam sebuah transaksi, tanggal dilakukannya transaksi.

3.9 Normalisasi Data

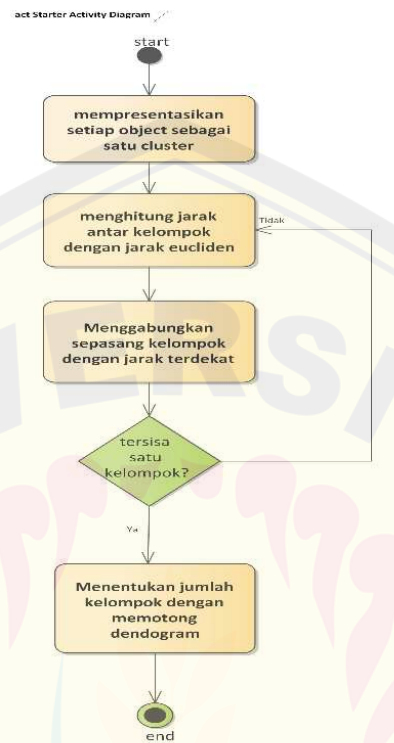
Proses normalisasi data pada penelitian menggunakan *Min-max* dimana hasil nilai *Min-max* merupakan hasil distribusi nilai baru yang memiliki mean sama dengan nol dan nilai deviasi standar sama dengan satu. Normalisasi pada model LRFM ini digunakan karena terdapat selisih nilai yang besar antara atribut L, R, F, M. Pada tahap ini menginisialisasi normalisasi *min-max* menggunakan fungsi *MinMaxScaler*. Selanjutnya pengolahan data bisa menggunakan fungsi *fit_transform()* dimana data yang sudah di *scaler* menggunakan fungsi *MinMaxScaler()* di sesuaikan dengan data dan di *transform* berdasarkan parameter data.

3.10 Clustering

Tahapan ini akan dilakukan dengan mengelompokkan data dengan data yang lain yang memiliki karakteristik yang sama. Metode yang digunakan dalam *clustering* ini adalah *Agglomerative hierarchical clustering*. Algoritma *Agglomerative hierarchical clustering* akan mengelompokkan data dengan jarak kedekatan antar data berdasarkan teknik jarak pada *Agglomerative hierarchical clustering*. Alur metode *Agglomerative Hierarchical clustering* menurut (Bachtiar, et.,al, 2017) seperti gambar 3.2.

Alur *Agglomerative hierarchical clustering* dimulai dari mempresentasikan data sebagai *cluster* kecil (*leaf*). Selanjutnya algoritma akan menghitung jarak kedekatan antar data menggunakan jarak *Euclidean distance*. Data yang di gambarkan sebagai *cluster* kecil (*leaf*) akan digabungkan berdasarkan jarak terkecil dibanding dengan jarak antar data lainnya. Kemudian dilakukan pembaruan menghitung jarak antar *clusternya* sesuai dengan metode jarak pada *Agglomerative hierarchical clustering* yaitu, *Single Linkage*, *Complete Linkage*, *centroid linkage*, *Average Linkage* dan *ward's linkage*. Pengelompokkan data tersebut akan diulangi

sampai *leaf* atau k sama dengan 0. Apabila $k = 0$ tahap terakhir adalah menentukan jumlah kelompok dengan memotong dendrogram dengan batas nilai.



Gambar 3.2 Alur Agglomerative hierarchical clustering

3.11 Pengujian Hasil *Clustering*

Pengujian nilai *cluster* dilakukan untuk mengetahui hasil nilai *cluster* terbaik pada proses *clustering* data. Proses pengujian *cluster* pada tahap ini menggunakan pengujian silhouette. Pengujian silhouette akan dilakukan terhadap semua nilai k (*cluster*) yang terbentuk pada hasil *clustering* menggunakan metode *Agglomerative hierarchical clustering*. Evaluasi hasil nilai k terbaik dilakukan pada nilai silhouette yang mendekati angka 1. Hal ini didukung oleh penelitian (Pramessti, Furqon, dan Dewi 2017) dimana jika nilai silhouette mendekati nilai 1 maka semakin baik pengelompokan data dalam suatu *cluster*. Jika nilai silhouette mendekati nilai -1 maka semakin buruk hasil *cluster* dalam suatu *cluster*.

3.12 Penentuan Nilai CLV

Untuk menentukan nilai setiap segmen pelanggan bisa menggunakan perhitungan CLV dengan persamaan 7. Dimana nilai $C_L^j, C_R^j, C_F^j, C_M^j$ merupakan nilai LRFM hasil segmentasi pelanggan berdasarkan hasil *cluster*, dan nilai w_L, w_R, w_F, w_M merupakan nilai hasil LRFM yang diperoleh dari AHP pada gambar 7. Nilai w_L, w_R, w_F, w_M dengan nilai w_L 0.238, w_R 0.88, w_F 0.326, w_M 0,3248 (Parvaneha, Abbasimehrb, dan Tarokhc 2012). Selanjutnya nilai Segmentasi berdasarkan CLV akan di urutkan berdasarkan nilai CLV yang terbesar ke yang terkecil.

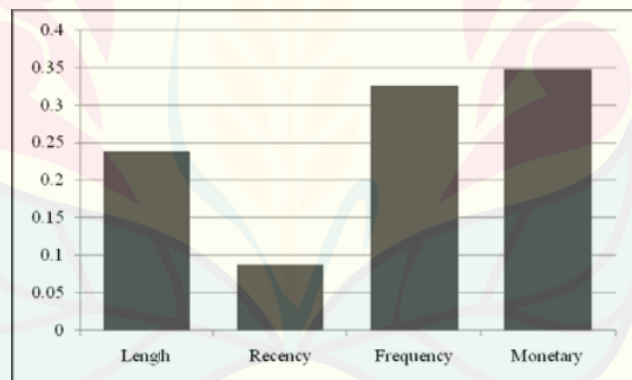
$$C^j = w_L C_L^j + w_R C_R^j + w_F C_F^j + w_M C_M^j \quad (7)$$

Keterangan:

C_j = Cluster j

w_L, w_R, w_F, w_M = Bobot variabel L, R, F, M

$C_L^j, C_R^j, C_F^j, C_M^j$ = Nilai rata-rata *cluster j* terhadap variabel L, R, F, M

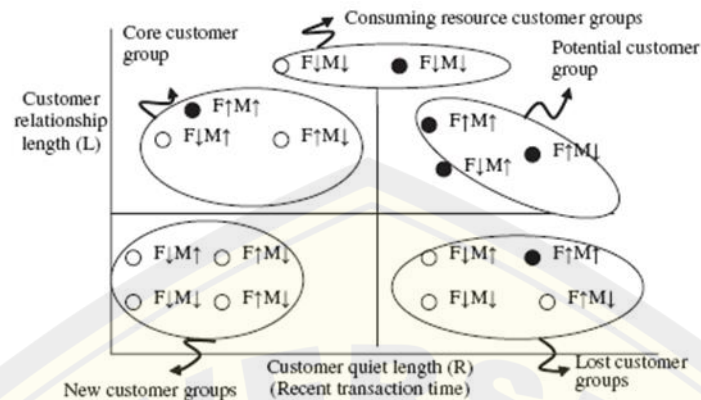


Gambar 3.3 Pembobotan LRFM menggunakan AHP oleh Parvaneha, Abbasimehrb, and Tarokhc 2012x

3.13 Analisa Dan Visualisasi Data

Dalam menentukan segmentasi pelanggan, (Zoeram dan Mazidi 2018) menggunakan *customer loyalty matrix* pada gambar 3.1, yang mewakili model LRFM untuk mencari segmen pelanggan dari kedua *matrix* tersebut didapatkan 16 kelompok segmen pelanggan dan 5 kelompok jenis pelanggan. Pengelompokan

clustering agar memudahkan membaca penjabaran jenis kelompok dan nama kelompok dapat dilihat pada tabel 1 (Zoeram dan Mazidi 2018).



Gambar 3.4 Customer Loyalty Matrix untuk Customer clustering mengacu nilai Length dan Recency

Tabel 3.1 Deskripsi kelompok dan cluster

No	Nama Group	Cluster	Nama Cluster	LRFM	Tipe Cluster
1	Core customer	C1	High value loyal customer	↑↓↑↑	LFM
		C2	Platinum customer	↑↓↑↑	LM
		C3	High Frequency buying customer	↑↓↑↓	LF
2	Potential Customer	C4	Potential loyal customer	↑↑↑↑	LRFM
		C5	Potensial consumption customer	↑↑↑↑	LRM
		C6	Potential high Frequency customer	↑↑↑↓	LRF
3	New Customer	C7	High value new customer	↓↓↑↑	FM

No	Nama Group	Cluster	Nama Cluster	LRFM	Tipe Cluster
		C8	<i>Spender promotion customer</i>	↓↓↓↑	M
		C9	<i>Frequency promotion customers</i>	↓↓↓↑	F
		C10	<i>Uncertain new customer</i>	↓↓↓↓	<i>Uncertain</i>
		C11	<i>High value lost customers</i>	↑↑↑↑	RFM
		C12	<i>Consumption lost customers</i>	↑↑↑↑	RM
4	<i>Lost Customer</i>	C13	<i>Frequency lost customers</i>	↑↑↑↓	RF
		C14	<i>Uncertain lost customers</i>	↑↑↓↓	R
		C15	<i>Low consumption cost customers</i>	↑↓↓↓	L
5	<i>Resource Customer</i>	C16	<i>High consumption cost customers</i>	↑↑↓↓	LR

Core Customer atau pelanggan inti merupakan pelanggan yang di kelompokkan sebagai pelanggan inti. Kelompok ini mencakup 3 *cluster* yaitu, C1, C2, C3. *Cluster C1 High value loyal customer* ini menggambarkan kelompok pelanggan yang memiliki nilai hubungan tinggi dengan perusahaan, dengan *Frequency* pembelian tinggi dan uang yang dikeluarkan pelanggan kepada perusahaan tinggi sehingga di gambarkan dengan LFM. *Cluster C2 Platinum customer* kelompok pelanggan yang memiliki nilai hubungan yang tinggi dengan perusahaan dan uang yang dikeluarkan pelanggan juga tinggi sehingga di gambarkan sebagai LM. *C3 High Frequency buying customer* kelompok pelanggan

yang memiliki nilai hubungan tinggi dan *Frequency* pembelian naik sehingga digambarkan dengan LF.

Potential Customer atau pelanggan potensial merupakan pelanggan yang dikelompokkan pada pelanggan yang diprediksi mempunyai kemungkinan untuk menjadi *core customer*. Kelompok *Potential Customer* terdiri dari 3 *cluster* yaitu C4, C5, C6. C4 *Potential loyal customer* ini menggambarkan kelompok pelanggan yang memiliki nilai hubungan dengan perusahaan yang tinggi, rentang waktu pembelian yang tinggi, *Frequency* pembelian yang tinggi dan jumlah uang yang dikeluarkan tinggi sehingga digambarkan dengan LRFM. C5 *Potensial consumption customer* menggambarkan kelompok pelanggan yang memiliki nilai hubungan dengan perusahaan tinggi, rentang waktu pembelian tinggi dan jumlah uang yang di keluarkan pelanggan tinggi digambarkan dengan LRM. C6 *Potential high Frequency customer* menggambarkan kelompok pelanggan yang memiliki nilai hubungan dengan perusahaan tinggi, rentang waktu pembelian tinggi dan, *Frequency* pembelian tinggi digambarkan dengan LRF.

New Customer atau kelompok pelanggan baru merupakan kategori pelanggan dengan nilai hubungan dengan perusahaan yang cenderung turun. Pada kelompok ini terdapat 4 kelompok *cluster* yaitu C7, C8, C9, C10. C7 *High value new customer* dengan nilai *Frequency* naik dan uang yang dikeluarkan tinggi digambarkan dengan FM. C8 *Spender promotion customer* dengan uang yang di keluarkan tinggi sehingga digambarkan dengan M. C9 *Frequency promotion customers* dengan nilai *Frequency* pembelian naik digambarkan dengan F. C10 *Uncertain new customer* dengan nilai hubungan dengan pelanggan rendah, rentang waktu pembelian rendah, *Frequency* rendah dan uang yang dikeluarkan dengan rendah. Sehingga digambarkan dengan uncertain.

Lost customer atau kelompok pelanggan lama namun hubungan dengan perusahaan cenderung menurun. Pada kelompok ini terdapat 4 *cluster* yaitu C11, C12, C13, C14. C11 *High value lost customer* karakteristik pada *cluster* ini adalah nilai *Recency*, *Frequency*, dan *spend Monetary* yang tinggi sehingga digambarkan dengan RFM. C12 *Consumption lost customer* dengan karakteristik nilai *Recency* dan *Monetary* yang tinggi sehingga digambarkan dengan RM. C13

Frequency lost customer dengan karakteristik nilai *Recency* dan *Frequency* yang tinggi sehingga digambarkan dengan RF. C14 *Uncertain lost customer* dengan karakteristik pelanggan *Recency* sehingga digambarkan dengan R.

Consuming resource customer atau kelompok pelanggan yang berdasarkan dengan sumber daya yang dikonsumsi pelanggan, pada kelompok ini terdapat 2 *cluster* yaitu C15, C16. C15 *low consumption lost customer* pelanggan di *cluster* ini digambarkan dengan hubungan perusahaan naik namun variabel lainnya menurun sehingga digambarkan dengan nilai L. C16 *high consumption lost customer* pelanggan dengan *cluster* ini memiliki atau digambarkan hubungan dengan perusahaan yang tinggi dan *Recency* yang tinggi sehingga digambarkan dengan LR.

3.14 Analisa Strategi CRM

Pada tahap ini akan dilakukan analisa perilaku pelanggan berdasarkan segmen yang dihasilkan. Dari analisa hasil *clustering* yang dilakukan akan dilakukan strategi promosi pada masing masing segmen. Strategi ini disusun untuk memberikan saran strategi CRM yang kepada KPRI UNEJ

BAB 4 HASIL DAN PEMBAHASAN

4.1 Data Collection

Tahapan pengumpulan data bertujuan untuk mendapatkan data recap yang dibutuhkan dalam penelitian. Data yang didapatkan dari KPRI UNEJ terdiri dari banyak variable data seperti tanggal, kode barang, nama barang, Isi, kode lokasi toko, keterangan nomor, kode anggota atau Id member, cara bayar hingga pajak atau tidak. Ada setidaknya 36 variabel dalam data transaksi penjualan pada KPRI UNEJ. Data yang digunakan pada penelitian ini adalah data transaksi penjualan KPRI UNEJ dari bulan November 2021 hingga April 2022. Jumlah transaksi pada 6 bulan tersebut terdapat 355,097 transaksi produk.

4.2 Data Preprocessing

Tahap data preprocessing menghilangkan variabel atau menghapus data kosong pada dataset penelitian. Pada tahap data collection sebelumnya didapatkan data berupa data transaksi penjualan sebanyak 355,097. Data transaksi tersebut belum dilakukan pembersihan dan masih memiliki data kosong, maka pada proses ini data kosong dihilangkan dengan memfilter variable kode anggota yang memiliki nomor kode anggota. Variable kode anggota yang memiliki angka 0 merupakan pelanggan umum dan data ini tidak digunakan karena pelanggan umum bersifat umum dan keseluruhan transaksi tanpa kode anggota.

Proses ini menghapus variable lain yang tidak digunakan. Variabel yang tidak digunakan dan keterangan variabel dapat dilihat pada tabel 4.1. Sedangkan variabel yang digunakan pada penelitian ini dapat dilihat pada Tabel 4.2. Hasil dari data processing didapatkan data transaksi sebanyak 120,066 data transaksi produk pada dan 17 atribut yang digunakan pada penelitian.

Tabel 4.1 Tabel variable yang dihapus dan tidak digunakan pada penelitian

Nama Atribut	Keterangan
Satuan	Menjelaskan Untuk Jenis Satuan Barang Yang Terjual

Nama Atribut	Keterangan
Discount	Discount Untuk Barang Satuan %
Total	Total Pembelian Anggota Pada Transaksi Tersebut
Potongan	Potongan Harga Yang Didapat Pelanggan Saat Melakukan Transaksi
Cara Bayar	Cara Bayar Pelanggan Secara Tunau Atau Melalui Mesin EDC
Sub Total	Total Pembelian Anggota Pada Transaksi Tersebut Setelah Dilakukan Pemotongan
Bayar	Jumlah Uang Yang Dibayarkan Oleh Pelanggan
Piutang	Jumlah Piutang Yang Dilakukan Oleh Pelanggan
Nama User	Nama Kasir Atau Petugas Yang Melayani Proses Transaksi
Kelompok	Menandakan Kode Pelanggan Dalam Kelompok
Kode Lokasi	Berisikan Kode Lokasi Toko
Pajak	Menandakan Apakah Transaksi Pelanggan Tersebut Pajak Atau Tidak

Tabel 4.2 Nama Atribut dan Penjelasan atribut yang digunakan

Atribut	Penjelasan
InvoiceNo	Nomer Invoice atau faktur tanda pembelian suatu produk
Tanggal	Tanggal atau waktu faktur tanda pembelian dilakukan/dibuat
StockCode	Kode stok barang yang telah terjual
Description	Dekripsi nama produk yang terjual
Quantity	Jumlah satuan barang yang terjual
Unitprice	Harga satuan produk yang terjual
Customer Id	Id pelanggan yang digunakan untuk pengenalan/ Nomer Member pelanggan
Country	Asal atau departemen pelanggan KPRI UNEJ

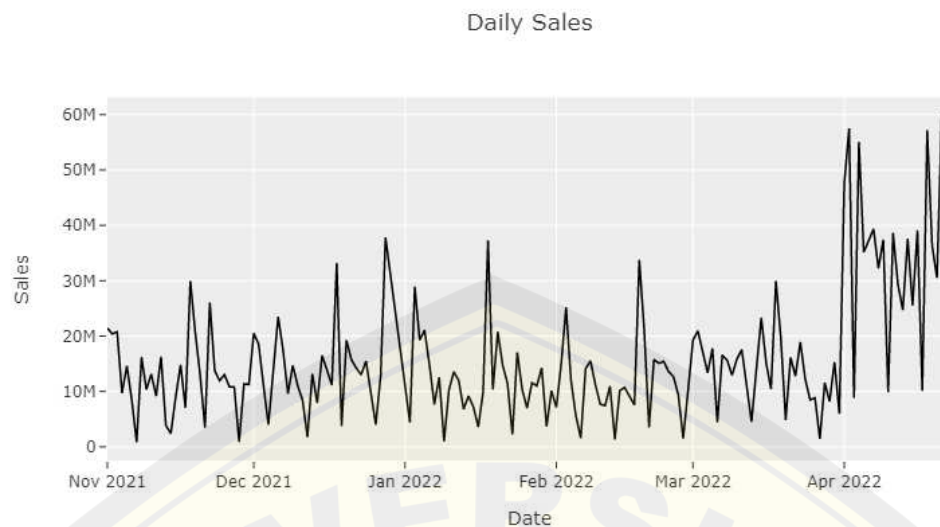
Atribut	Penjelasan
Jam	Jam faktur tanda pembelian dilakukan
InvoiceDate	Tanggal dan waktu faktur tanda pembelian dilakukan
Year	Tahun faktur tanda pembelian dilakukan
Month	Bulan faktur tanda pembelian dilakukan
Week	Minggu faktur tanda pembelian dilakukan
Year month	Tahun dan bulan faktur tanda pembelian dilakukan
jam pembulatan	Jam faktur tanda pembelian dilakukan /tapi lebih dibulatkan lagi
is_canceled	Transaksi yang dilakukan mengalami retur atau pengembalian
Date	Tanggal, Bulan, dan Tahun faktur pembelian dilakukan/dibuat

4.3 Data Understanding

Data understanding merupakan tahapan dimana melakukan pemahaman pada data sebelum diolah lebih lanjut. Tahap ini dilakukan analisa terkait penjualan harian yang dikemas dalam data selama periode November 2021 hingga April 2022, analisa terkait total transaksi selama perbulan, dan analisa seberapa besar transaksi yang dilakukan selama perjam dan hari apa transaksi paling tinggi dilakukan. Berikut penjelasan lebih lanjut mengenai grafik yang dianalisa.

4.3.1 Grafik Penjualan Harian

Data grafik penjualan harian dilakukan untuk mengetahui penjualan selama sehari dalam periode bulan November 2021 hingga April 2022. Berdasarkan atribut pada Tabel 4.2 atribut yang digunakan adalah atribut Date. Angka penjualan dikelompokkan berdasarkan setiap tanggal penjualan dan total penjualan setiap hari yang didapat dari Quantity pertransaksi dikali harga perunit produk. Dari hasil tersebut didapatkan grafik seperti gambar 4.1.

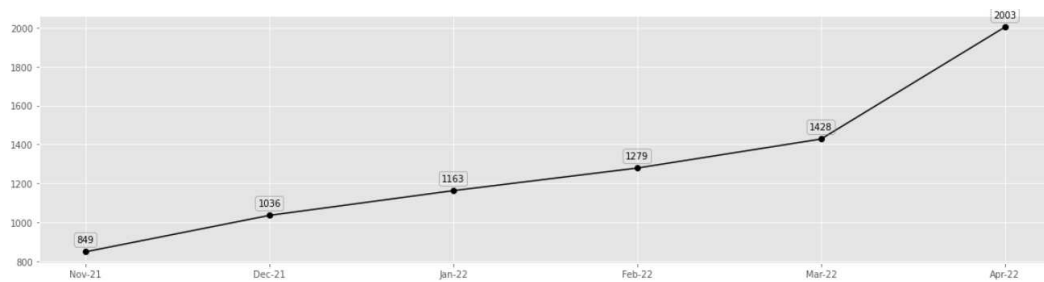


Gambar 4.1 Grafik Garis Penjualan Harian

Berdasarkan gambar diatas dapat diketahui bahwa dalam penjualan perhari dari bulan November 2021 hingga Maret 2022 penjualan tidak bisa dikatakan stabil karena pada 5 bulan tersebut bentuk grafik naik turun tidak menentu. Penjualan mengalami kenaikan pada bulan April 2022 dimana penjualan tertinggi berada tanggal 21 bulan April dengan total penjualan sebanyak Rp. 5.982.965 juta dan penjualan terendah berada di tanggal 7 November 2021 dengan total penjualan sebanyak Rp. 812.900 ribu

4.3.2 Grafik Jumlah Customer Perbulan

Data grafik jumlah customer penjualan menganalisa jumlah pelanggan dari dari bulan ke bulan dalam periode bulan November 2021 hingga April 2022. Atribut yang digunakan perbandingan pada gambar 4.2 adalah bulan selama periode penelitian atau date dan jumlah pelanggan. Nilai jumlah pelanggan didapat dari menjumlahkan kunjungan yang dilakukan pelanggan pada hari tersebut.

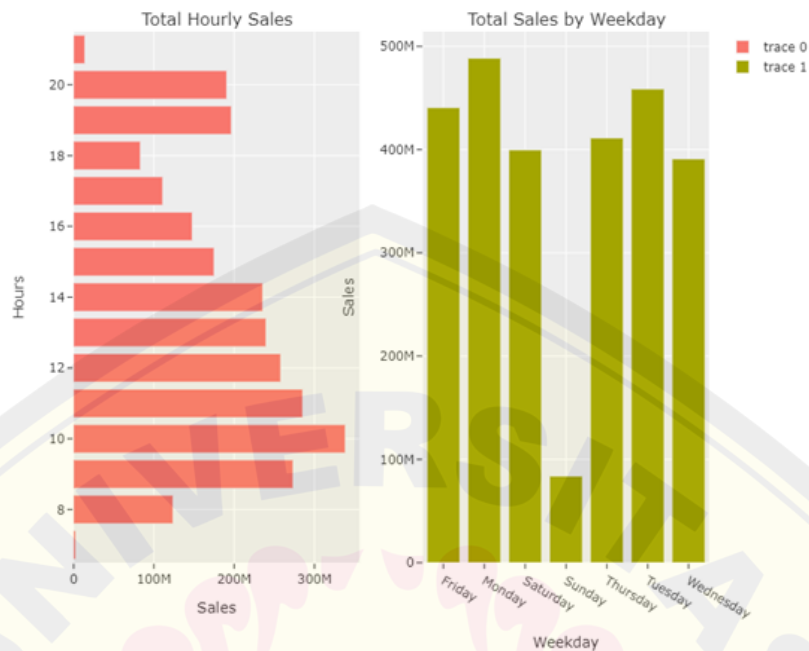


Gambar 4.2 Grafik Penjualan Perbulan

Berdasarkan gambar 4.2 dapat diketahui bahwa setiap bulanya pengunjung atau pembeli di toko KPRI UNEJ mengalami peningkatan. Hal ini dapat dilihat dari peningkatan dari bulan ke bulan dapat dilihat pada gambar 4.2 dimana pada bulan November terdapat kunjungan sebanyak 849 pelanggan dan pada bulan Desember terdapat kunjungan sebanyak 1036, Januari sebanyak 1163 pelanggan, Februari sebanyak 1279 pelanggan, Maret sebanyak 1428 pelanggan, dan April sebanyak 2003 pelanggan. Peningkatan tertinggi terjadi pada bulan April 2022 dimana pada bulan tersebut toko KPRI UNEJ mengalami total kunjungan pelanggan sebanyak 2003 pelanggan.

4.3.3 Grafik Penjualan Per Jam dan Penjualan Per Hari

Data grafik penjualan per jam dan penjualan perhari menganalisa jumlah rata rata penjualan per jam dan per harinya. Atribut yang digunakan perbandingan pada gambar 4.3 adalah jam penjualan, weekday.



Gambar 4.3 Grafik penjualan per jam dan per hari

Berdasarkan gambar 4.3 dapat diketahui bahwa jam buka toko dimulai dari jam 8 pagi hingga 9 malam. Pada jam kerja tersebut dapat diketahui bahwa penjualan tertinggi terjadi pada pukul 10 pagi dengan total penjualan sebanyak Rp. 3.382.633 juta, penjualan paling sedikit terjadi pukul 7 dimana jam tersebut merupakan jam awal buka toko. Bisa dikatakan bahwa toko mungkin belum stabil. Selain itu jam penjualan ramai lain yaitu pada jam 11 siang dimana penjualan pada jam tersebut berada di angka Rp. 2.851.701.

Berdasarkan gambar 4.3 dapat diketahui bahwa penjualan tertinggi terjadi di hari senin dengan total penjualan sebanyak Rp. 4.884.338, lalu pada hari Jum'at penjualan sebanyak Rp. 4.585.357. Jadi dapat disimpulkan bahwa hari paling padat penjualan berdasarkan data transaksi mulai dari November 2021 hingga April 2022 terjadi pada hari Senin..

4.4 Pembentukan Model LRFM

Pembentukan atribut model LRFM diperoleh dari pengolahan data collection pada tabel 4.2. Atribut yang akan digunakan untuk membentuk model LRFM ada lima diantaranya atribut CustomerId, InvoiceNo, Sales, InvoiceDate, Date.

Nilai variable *Length* atau L merupakan selisih antara tanggal pertama melakukan transaksi pembelian dan tanggal terakhir melakukan transaksi pembelian. Atribut yang digunakan untuk membentuk variable *length* adalah atribut CustomerId dan Date. Atribut CustomerId digunakan untuk mengelompokkan seluruh data Date berdasarkan CustomerId.

Nilai variable *Recency* atau R merupakan selisih waktu dari tanggal pelanggan melakukan transaksi saat ini dengan tanggal yang digunakan acuan dalam penelitian. Tanggal acuan didapat dari tanggal minimum pada atribut InvoiceDate yaitu tanggal 1 November 2021. Pembentukan variable R yang merupakan kumpulan data dari InvoiceDate setiap CustomerId. Hasil tersebut akan dilakukan penghitungan selisih antara tanggal transaksi terbaru dikurangi dengan transaksi minimum acuan.

Nilai variable *Frequency* atau F merupakan jumlah transaksi selama November 2021 hingga April 2022. Atribut yang digunakan untuk membentuk variable ini adalah CustomerId dan InvoiceNo. Variable F didapat dengan menjumlah seluruh nomor faktur pembelian atau atribut InvoiceNo pembeli.

Nilai variable *Monetary* atau M merupakan jumlah uang yang dikeluarkan pelanggan pada semua transaksinya selama November 2021 hingga April 2022. Atribut yang digunakan untuk membentuk variable ini adalah Quantity, UnitPrice, CustomerId. Variabel M didapat dari mengalikan jumlah barang dan harga per barang kemudian total penjualan tersebut di kelompokkan terhadap konsumen.

Sehingga melalui algoritma yang dilakukan maka diperoleh 2003 baris data yang berisikan customer id pelanggan serta nilai variable untuk model LRFM. Contoh data dapat dilihat pada Tabel 4.3. Algoritma yang digunakan pada pembuatan model LRFM seperti pada gambar 4.4.

```

LRFM = data_.groupby('CustomerId').agg(Frequency=pd.NamedAgg(column="InvoiceNo", aggfunc="nunique")
, Monetary=pd.NamedAgg(column="Sales", aggfunc="sum"),
Recency = pd.NamedAgg(column='InvoiceDate', aggfunc='min')).reset_index()
length = data_.groupby('CustomerId')['Date'].max() - data_.groupby('CustomerId')['Date'].min()
length = (length/np.timedelta64(1, 'D')).reset_index()
length.columns = ['CustomerId', 'Length_of_stay']

LRFM = LRFM.merge(length, on='CustomerId', how='inner')
del length

LRFM.head(100)

```

Gambar 4.4 Code Pemodelan Model LRFM

Berdasarkan hasil *running Code* pada gambar 4.4 dihasilkan nilai seperti pada Tabel 4.3. Hasil meliputi nomer *customerId*, *Length*, *Recency*, *Frequency* dan *Monetary*.

Tabel 4.3 Hasil Pemodelan Model LRFM

No	CustomerId	Length_of_stay	Recency	Frequency	Monetary
1	0002	2.0	137	38	3.21E+06
2	0004	0.0	103	2	4.75E+04
3	0013	171.0	1	10	1.24E+06
4	0046	49.0	94	1	1.98E+05
5	0083	165.0	5	5	1.35E+06
6	0118	0.0	171	3	3.82E+05
7	0155	33.0	131	3	4.92E+05
8	0253	168.0	1	1	2.86E+05
9	0260	123.0	31	2	8.00E+03
10	0269	0.0	1	6	1.46E+06

4.5 Normalisasi

Normalisasi data pada tahap penelitian ini bertujuan untuk membuat nilai dari masing-masing variable dalam rasio nilai yang sama atau setara. Proses normalisasi dilakukan dengan menggunakan fungsi dari python. Fungsi yang digunakan yaitu *MinMaxScaler* dan *fit transform*. Tahap pertama pada proses normalisasi adalah dengan membuat variable baru yaitu *min_max_scaler* dimana variable ini akan menyimpan hasil inisialisasi normalisasi *MinMaxScaler*. Selanjutnya membuat variable baru yang digunakan untuk menyimpan hasil

normalisasi LRFM yaitu variable LRFM_2. Code normalisasi dapat dilihat pada gambar 4.5.

```
from sklearn import preprocessing

min_max_scaler = preprocessing.MinMaxScaler(feature_range=(0,1)) #inisialisasi normalisasi MinMax
LRFM_2 = min_max_scaler.fit_transform(LRFM_1) #transformasi MinMax untuk fitur
#LRFM_2 = pd.DataFrame({'Frequency':array[:,0], 'Monetary':array[:,1], 'Recency':array[:,2], 'Lenght_of_stay':array[:,3]})

print("dataset setelah dinormalisasi :")
print(LRFM_2)
```

Gambar 4.5 Normalisasi Menggunakan MinMax

Hasil normalisasi data masih berbentuk array diubah dalam bentuk dataframe menggunakan pandas pd.DataFrame dan juga menamai untuk header penamaan nama kolom pada model LRFM. Data yang dirubah dalam bentuk dataframe dapat dilihat pada Tabel 4.4.

Tabel 4. 4 Hasil Normalisasi Model LRFM Menggunakan MinMax scaler

<i>Length_of_stay</i>	<i>Recency</i>	<i>Frequency</i>	<i>Monetary</i>
0.01	0.79	0.01	0
0	0.6	0	0
0.99	0.01	0.21	0.05
0.28	0.54	0.01	0
0.95	0.03	0.05	0.02
0	0.99	0	0
0.19	0.76	0.02	0.02
0.97	0.01	0.01	0.01
0.71	0.18	0.01	0.01
0	1	0	0
0.12	0.43	0.01	0

4.6 Clustering

4.6.1 Clustering Menggunakan Agglomerative Hierarchical Clustering

Secara garis besar proses *clustering* menggunakan metode *Agglomerative hierarchical clustering* untuk pengelompokkan data pada penelitian ini adalah dengan menggunakan library *scipy.cluster.hierarchy* as *sch* dan *sklearn.cluster*

import AgglomerativeClustering. Library scipy sendiri merupakan perpustakaan *opensource* yang digunakan untuk perhitungan ilmiah tingkat tinggi. Sedangkan Modul Sklearn.Cluster sendiri adalah membantu processing data ataupun training data untuk kebutuhan *machine learning*. Banyak fitur yang ada pada modul ini seperti model klasifikasi, clustering, regresi, mahine learning. Pada tahap ini dilakukan percobaan macam-macam linkage yang ada pada metode Agglomerative Hierarchical clustering diantaranya ada *Single Linkage*, *Complete Linkage*, *Average Linkage* dan Ward Linkage. Fungsi yang digunakan pada library ini adalah dendrogram, method='ward', 'single', 'complete', dan 'Average' untuk code lebih lanjut dapat dilihat pada gambar 4.6. Penggunaan masing-masing percobaan cara linkage pada code dapat mengganti method sesuai dengan linkage yang digunakan.

```
plt.figure(figsize=(10, 7))
plt.title("Dendrograms")
dend = sch.dendrogram(sch.linkage(LRFM_3, method='ward'))
```

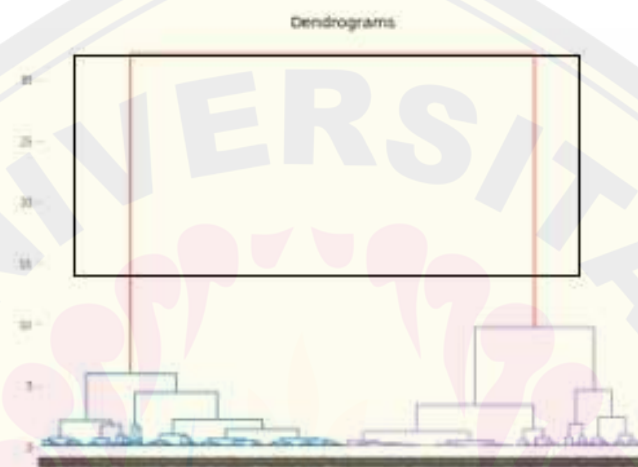
Gambar 4.6 Code Clustering dan Plot Menggunakan Dendogram

Alur untuk proses cluster ini akan membandingkan macam-macam linkage dan linkage yang terbaik akan digunakan sebagai jarak atau linkage clustering pada data penelitian ini. Pengukuran kedekatan dilakukan pada data dan didapatkan gambar dendogram seperti pada lampiran. Dimana dari gambar dapat diketahui bahwa ward linkage, *Single Linkage*, *Complete Linkage* memiliki bentuk dendogram yang teratur sedangkan *Single Linkage* memiliki bentuk dendogram yang tidak beraturan.

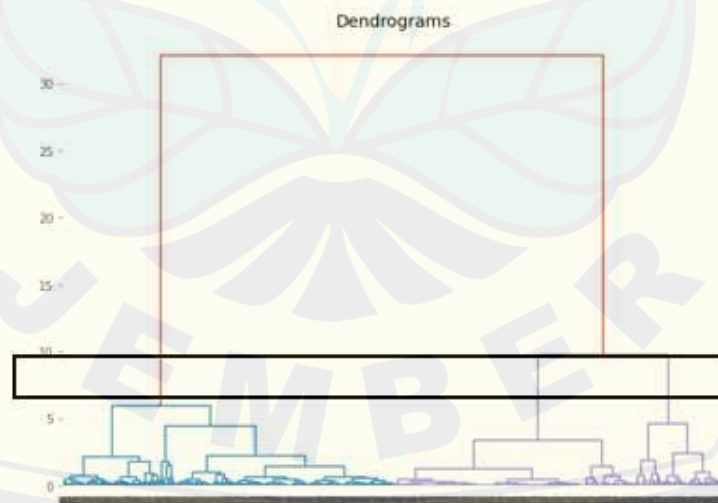
4.6.1 Menentukan *cluster* dengan melakukan pemotongan dendogram

Hasil *clustering* pada penelitian menggunakan algoritma Agglomerative Hierarcical *clustering* dapat dilihat menghasilkan ppohon dendogram seperti pada gambar. Berdasarkan gambar tersebut dapat diketahui bahwa secara garis besar hasil *clustering* mengelompokkan data pelanggan KPRI

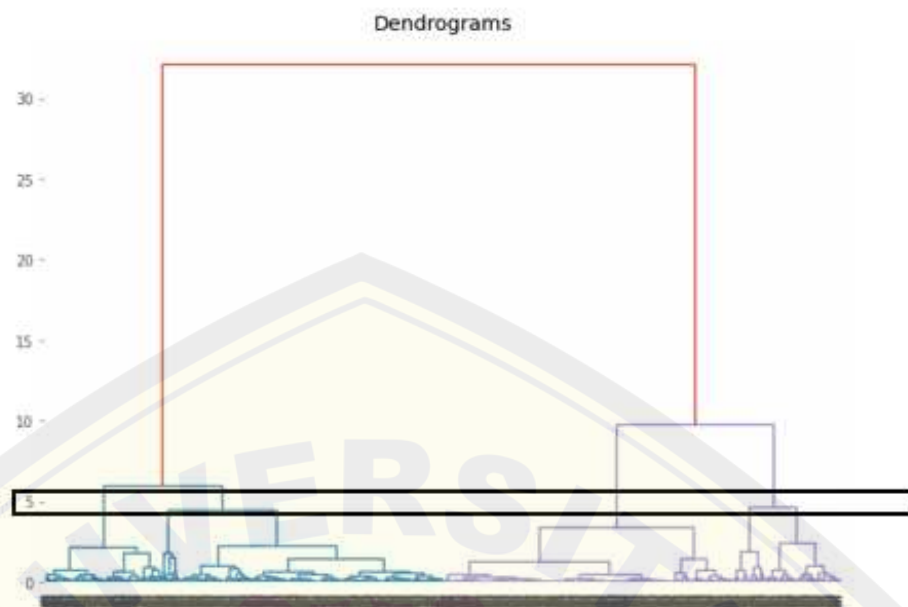
UNEJ menjadi 2 kelompok pelanggan. Namun untuk mengetahui lebih lanjut *cluster* terbaik dapat dilakukan pemotongan dendogram. pemotongan dendogram dilakukan dengan mengamati dendogram. misal pada gambar 4.7 diketahui bahwa *cluster* terbaik yang terbentuk adalah 2. Pemotongan dilakukan dengan membuat garis yang tidak melewati garis vertical horizontal dendogram. untuk pemotongan lainnya dapat dilihat pada gambar gambar 4.8, gambar 4.9 dan gambar 4.10.



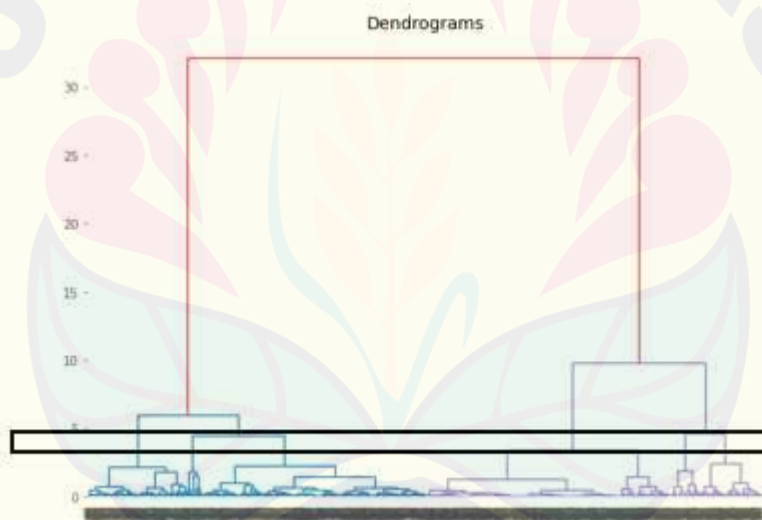
Gambar 4.7 Gambar Hasil Clustering Menggunakan AHC Menggunakan Dendogram



Gambar 4.8 Gambar pemotongan dendogram untuk 3 cluster



Gambar 4.9 Gambar pemotongan dendrogram untuk 4 cluster



Gambar 4.10 Gambar pemotongan dendrogram untuk 6 cluster

4.7 Pengujian *cluster*

Pengujian *cluster* dilakukan dengan menggunakan silhouette score dimana nilai jumlah kluster didapat dari pemotongan pada gambar dendrogram pada masing masing jenis linkage untuk menemukan jumlah *cluster* yang terbaik. Nilai *cluster* yang digunakan pada penelitian ini adalah 2,3,4, dan 6. Pengujian dilakukan dengan menggunakan code python seperti pada gambar 4.11.

```

# Silhouette analysis
range_n_clusters = [2,3,4,6]

for num_clusters in range_n_clusters:

    hc = AgglomerativeClustering(n_clusters = num_clusters , affinity = 'euclidean', linkage = 'ward')
    y_hc = hc.fit_predict(LRFM_3)
    cluster_labels = hc.labels_

    # silhouette score
    silhouette_avg = silhouette_score(LRFM_3, cluster_labels)
    print("For n_clusters={0}, the silhouette score is {1}".format(num_clusters, silhouette_avg))

```

Gambar 4.11 Code Pengujian Hasil Cluster

Berdasarkan hasil running code python pada gambar 4.11 dihasilkan perbandingan antara jumlah *cluster* dengan nilai silhouette pada masing-masing linkage yang dapat dilihat pada tabel 4.5.

Tabel 4.5 Hasil Perhitungan Silhouette terhadap kemungkinan jumlah cluster

Linkage	Nilai Silhouette pada Cluster			
	2	3	4	6
Ward	0.71667	0.65000	0.53805	0.57742
Single	0.17200	0.17076	0.16533	-0.14460
Average	0.70206	0.68848	0.63663	0.58431
Complete	0.60466	0.57553	0.61415	0.60452

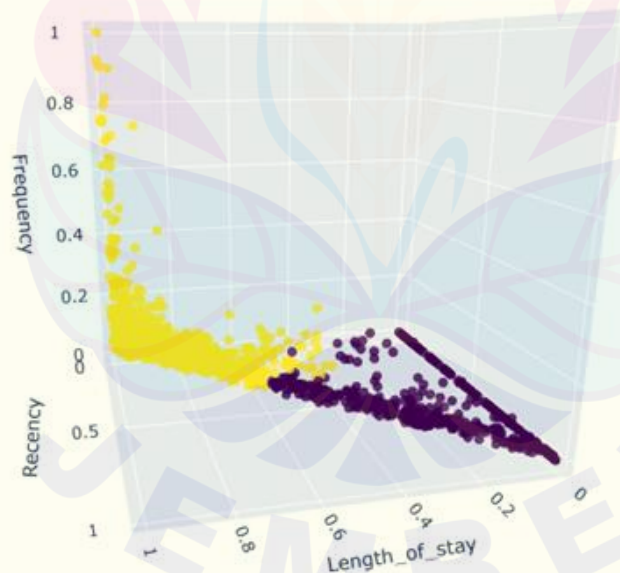
Berdasarkan hasil penghitungan tersebut dapat dilihat pada tabel 4.5 bahwa nilai silhouette tertinggi berada diangka k=2 dengan nilai silhouette tertinggi 0.71667 dimana linkage terbaik adalah ward linkage. nilai silhouette semakin mendekati angka 1 maka dapat dikatakan bahwa semakin baik pengelompokkan sebuah *cluster* (Pramesti, Furqon, dan Dewi 2017). Dari Tabel 4.5-1 dapat diketahui bahwa nilai silhouette yang mendekati angka 1 adalah k=2 dengan jarak linkage yang digunakan adalah ward linkage nilai silhouttenya adalah 0.716667. Maka

cluster yang akan digunakan untuk penelitian ini adalah 2 *cluster* dengan jarak ward linkage.

4.8 Visualisasi *cluster*

4.8.1 Visualisasi Data Dengan Menggunakan 3D

Grafik scatter plot 3D menunjukkan persebaran nilai objek pada model LRFM. Grafik 3D terdiri dari 3 sumbu yaitu x,y,z. Karena 3D terbatas oleh 3 variable maka dilakukan pemecahan terhadap model LRFM dimana model dipecah menjadi LRM, LRF, LFM dan RFM. Hasil pemecahan tersebut kemudian dijadikan perbandingan grafik dimana LRM merupakan perbandingan dari variabel *Length*, *Recency*, *Monetary*, LRF perbandingan dari variable *Length*, *Recency*, *Frequency*, LFM perbandingan dari variable *Length*, *Frequency*, *Monetary* dan terakhir RFM merupakan perbandingan dari variable *Recency*, *Frequency*, *Monetary*. Hasil gambar untuk perbandingan LRF dapat dilihat pada gambar 4.12. Hasil model LRM, LFM dan RFM dapat dilihat pada lampiran.

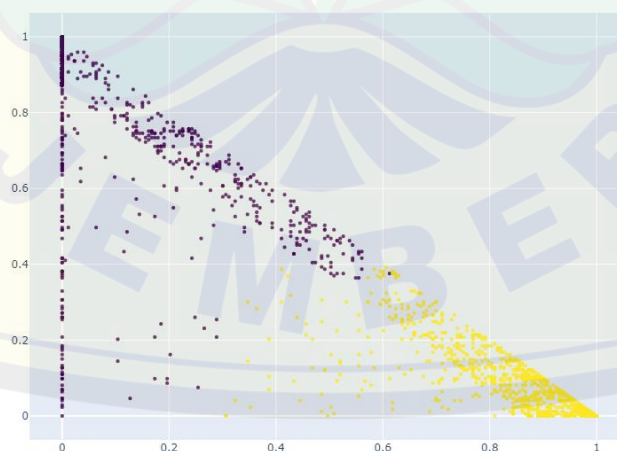


Gambar 4.12 Gambar Hasil Scatter plot 3D model LRF dimana nilai x = *Length_of_stay*, y = *Recency*, z = *Frequency*

Hasil scatter plot 3D pada gambar 4.12 dapat diketahui bahwa persebaran datanya terdapat perbedaan yang begitu jelas. Hal ini dapat dilihat dari anggota atau elemen data pada setiap *cluster* yang saling berdekatan. Hanya ada beberapa anggota *cluster* yang tidak berada pada *cluster* yang lainnya. Hal ini dapat disebabkan karena terdapat outlier data pada suatu data. Namun seperti yang terlihat data tidak terdapat data dengan karakteristik yang jauh berbeda dengan kelompok data lainnya. Hal ini bisa disebabkan karena data anggota *cluster* sudah dinormalisasi sehingga anggota data dalam suatu *cluster* masih memiliki karakteristik yang sama dengan anggota *cluster* lainnya.

4.8.2 Visualisasi Data Menggunakan 2D

Grafik 2D menggambarkan persebaran data / object pada model LRFM. Penggambaran 2D ini dilakukan untuk mengetahui perbandingan 2 variabel terhadap model LRFM. Grafik scatter plot 2D terdiri dari 2 dimensi yaitu x dan y. sedangkan model LRFM sendiri terdiri dari 4 variabel yaitu *length*, *Recency*, *Frequency*, *Monetary* sehingga untuk penggambaran grafik scatter plot 2d model LRFM harus dipecah. Pemecahan LRFM ini terdiri dari LR, LF, LM, RF, RM, dan FM. Dimana pemecahan tersebut akan dijadikan sebagai perbandingan pada scatter plot 2D. Hasil scatter plot LR yaitu perbandingan dari variable *length* dan *Recency* dapat dilihat pada gambar 4.13. Hasil perbandingan untuk variable lain dapat dilihat pada lampiran.



Gambar 4.13 Gambar Hasil Scatter plot 2D model LR. dimana nilai x = length, y = recency

Hasil scatterplot 2D pada gambar 4.13 dapat disimpulkan data menyebar dari sekitar garis diagonal dan mengikuti arah garis diagonal lalu menyebar kebawah. Dari gambar tersebut dapat diketahui juga bahwa persebaran data memiliki korelasi negative dimana artinya jika variabel x mengalami peningkatan maka variabel y akan mengalami penurunan dan jika variabel x mengalami penurunan maka variabel y akan mengalami peningkatan. Namun terdapat data yang tersebar diluar garis korelasi. Data tersebut dapat dikatakan sebuah outlier pada data penelitian. Hal ini dikarenakan pada variabel *length* dan *Recency* terdapat nilai 0 atau nilai data berada di luar garis korelasi hubungan negative.

4.9 Perhitungan atau pembobotan clv untuk masing masing cluster

Hasil *cluster* yang terbentuk pada tahap *clustering* akan dilakukan pembobotan untuk mengetahui nilai yang diberikan pelanggan kepada KPRI UNEJ. Pembobotan CLV dilakukan dengan mengalikan masing-masing variabel dengan nilai seperti pada persamaan 7. Penghitungan ini dilakukan dengan menggunakan code python seperti pada gambar

```
#PR OBJECT TO ARRAY
blok1 = sk.values
blok2 = [0.238,0.088,0.326,0.348] #urutannya dibalik
nilaiclv1 = []
print (blok1.dtype)

for x in blok1:
    nilaiclv =[]
    for z,i in enumerate(x):
        nilaiclv.append(i*blok2[z])
    nilaiclv1.append(nilaiclv)
print (nilaiclv1)
```

Gambar 4.14 Code Menentukan Nilai CLV untuk masing-masing cluster

Berdasarkan hasil running program tersebut ditemukan nilai seperti pada tabel 4.6

Tabel 4. 6 Hasil perkalian pembobotan dengan nilai ahp pada penelitian terdahulu

No	Nama <i>cluster</i>	L	R	F	M
1	<i>Cluster 0</i>	0.00163	0.00041	0.25474	0.03199
2	<i>Cluster 1</i>	0.00871	0.0067	0.03064	0.29082

4.10 Penentuan kelompok atau golongan *cluster*

Dari proses penghitungan nilai pembobotan CLV pada hasil *cluster* mendapatkan hasil berupa proses segmentasi pelanggan pada toko KPRI UNEJ, namun harus menentukan kriteria kelompok pelanggan. Penentuan ini dilihat dari rata-rata pada masing-masing variabel. Jika nilai variabel LRFM pada suatu *cluster* lebih tinggi dari rata-rata maka dapat dikatakan bahwa simbol LRFM yaitu $L\uparrow R\uparrow F\uparrow M\uparrow$. Penentuan simbol LRFM pada hasil *clustering* dapat dilihat pada Tabel 4.7.

Tabel 4.7 Tabel Hasil clustering berisikan rata-rata digunakan untuk menentukan simbol LRFM

No	Nama <i>cluster</i>	L	R	F	M
1	<i>Cluster 0</i>	0.00163	0.00041	0.25474	0.03199
2	<i>Cluster 1</i>	0.00871	0.0067	0.03064	0.29082
	Rata Rata	0.00517	0.003558	0.142688	0.161405

Dari Tabel di atas dapat disimpulkan bahwa *cluster 0* memiliki simbol $L\downarrow, R\downarrow, F\uparrow, M\downarrow$ dimana simbol tersebut berdasarkan Tabel 1. Dikatakan bahwa pelanggan tergolong sebagai *new customer* dengan nama *cluster Frequency promotion customers*. Sedangkan untuk *cluster 1* memiliki simbol dimana $L\uparrow, R\uparrow, F\downarrow, M\uparrow$ dikatakan bahwa pelanggan tergolong sebagai *potensial customer* dengan nama *cluster potensial consumption customer*. Jadi hasil segmentasi pelanggan KPRI UNEJ menghasilkan 2 *cluster*, *cluster 0* tergolong sebagai *new customer*,

cluster Frequency promotion customers dan *cluster 1* tergolong sebagai *potensial customer, potensial consumption customer*.

4.11 Analisa Strategi CRM Terhadap Hasil Cluster

4.11.1 Analisa hasil cluster 0

Cluster 0 memiliki jumlah anggota pelanggan sebanyak 998 pelanggan. dengan nilai CLV sebesar 0.288768 lebih rendah dari *cluster 1*. *Cluster 0* dikategorikan sebagai *new customer* dimana nilai Maka pelanggan termasuk kedalam jenis pelanggan *new customer* dengan nama pelanggan berdasarkan Tabel 1 termasuk kedalam *Frequency promotion customers*. Anggota *cluster 0* atau *new customer* dapat dilihat pada lampiran. Karakteristik pada *cluster* dapat dilihat pada tabel 4.8.

Tabel 4.8 Karakteristik Cluster 0 atau New Customer

	Length_of_stay	Recency	Frequency	Monetary
Min	0	0	1	Rp 8,000
Max	173	173	164	Rp 26,518,400
Rata-Rata	79	77	8	Rp 1,147,907

Kategori pelanggan *new customer* jumlah uang yang dikeluarkan kategori *cluster* ini sebanyak Rp. 1.145.611.548. Sedangkan rata-rata hubungan perusahaan atau *length* dengan pelanggan adalah 79 hari dengan nilai terendah 0 hari dan tertinggi 173 hari. Rata-rata *recency* pelanggan *new customer* atau jarak kunjungan pelanggan terhadap kunjungan terbaru adalah 77 hari dengan nilai terendah 0 hari dan tertinggi 173 hari. Rata-rata *frequency* kunjungan pelanggan sebanyak 8 kali dengan nilai terendah 1 kali dan tertinggi 164 hari. Sedangkan untuk variabel *Monetary* dengan rata-rata uang yang diterima dari kelompok pelanggan *new customer* sebanyak Rp. 1.147.907 dengan nilai terendah uang yang diterima Rp 8.000 dan nilai tertinggi sebesar Rp. 26.518.907.

Strategi CRM yang tepat menurut (Alizadeh Zoeram dan Karimi Mazidi 2018) terkait pelanggan dengan kategori *new customer* disarankan perusahaan untuk menaruh lebih banyak perhatian. Salah satu contohnya adalah memberi *transactional satisfaction* kepada pelanggan dengan kategori *new customer*. *Transactional satisfaction* atau kepuasan pelanggan bukan hanya terhadap layanan yang baik saja namun kepuasan transaksi spesifik dan keseluruhan kepuasan pelanggan dalam melakukan transaksi akan memicu perilaku pembelian kembali terhadap perusahaan (Bodet 2008).

Beberapa hal yang bisa dilakukan perusahaan untuk mengetahui *Transactional satisfaction* atau kepuasan pelanggan dapat dilakukan survei kepuasan dengan mengumpulkan *feedback* dari konsumen, memantau kompetitor perusahaan serta mempermudah akses konsumen saat melakukan belanja. Survei kepuasan pelanggan dilakukan untuk mengevaluasi kembali strategi pasti apa yang bisa dilakukan. Sebagai contoh seorang pelanggan mengatakan bahwa pelayanan pada suatu perusahaan dikatakan buruk maka perusahaan perlu memperbaiki kinerja layanan mereka. Berdasarkan hasil analisa grafik penjualan per jam dan per harian (dapat dilihat pada gambar 1.1) diketahui bahwa jam dengan rata-rata penjualan tertinggi ada pada jam 10. Sehingga survei pelanggan dapat dilakukan pada pukul tersebut. Rangkuman dari analisa hasil *cluster 0* dapat dilihat pada Tabel 8.

Tabel 4.9 Hasil Ringkasan analisa cluster 0

Cluster 0	
Nama Kategori Cluster	<i>New customer</i>
Peringkat Cluster	2
Jumlah Anggota	998
Nama Cluster	<i>Frequency promotion customers</i>
Strategi	Meningkatkan kepuasan pelanggan

Survei kepuasan pelanggan

Evaluasi hasil survei

Meningkatkan pelayanan

4.11.2 Analisa hasil *cluster* 1

Cluster 1 memiliki jumlah anggota pelanggan sebanyak 1005 pelanggan. dengan nilai CLV sebesar 0.336871 lebih tinggi dari *cluster* 0. *Cluster* 1 dikategorikan sebagai potensial customer dimana nilai *length* atau periode hubungan dengan perusahaan dikatakan naik atau cenderung meningkat, *Recency* atau selisih kunjungan pelanggan terhadap perusahaan dikatakan naik atau cenderung meningkat, *Monetary* atau uang yang dikeluarkan pelanggan untuk melakukan pembelian di toko juga dikategorikan naik atau cenderung meningkat, sedangkan untuk hasil *Frequency* pada *cluster* 1 *Frequency* atau waktu kunjungan pelanggan terhadap perusahaan rendah. Maka berdasarkan Tabel 1 pelanggan termasuk kedalam kategori pelanggan *potensial customer* dengan nama pelanggan *potensial consumption customer*. Anggota cluster 1 atau Potensial dapat dilihat pada lampiran. Karakteristik pada cluster dapat dilihat pada tabel 4.10.

Tabel 4.10 Karakteristik Cluster 0 atau Potensial Customer

	Length_of_stay	Recency	Frequency	Monetary
Min	0	0	1	Rp 15,000
Max	173	173	181	Rp 60,934,200
Rata-Rata	82	74	8	Rp 1,519,691

Kategori pelanggan *potensial customer* jumlah uang yang dikeluarkan kategori cluster ini sebanyak Rp. 1.527.289.900. Sedangkan rata-rata hubungan perusahaan atau *length* dengan pelanggan adalah 82 hari dengan nilai terendah 0 hari dan tertinggi 173 hari. Rata-rata *recency* pelanggan *new customer* atau jarak

kunjungan pelanggan terhadap kunjungan terbaru adalah 74 hari dengan nilai terendah 0 hari dan tertinggi 173 hari. Rata-rata *frequency* kunjungan pelanggan sebanyak 8 kali dengan nilai terendah 1 kali dan tertinggi 181 kali. Sedangkan untuk variabel Monetary dengan rata-rata uang yang diterima dari kelompok pelanggan new customer sebanyak Rp. 1.519.691 dengan nilai terendah uang yang diterima Rp 15.000 dan nilai tertinggi sebesar Rp. 60.934.200.

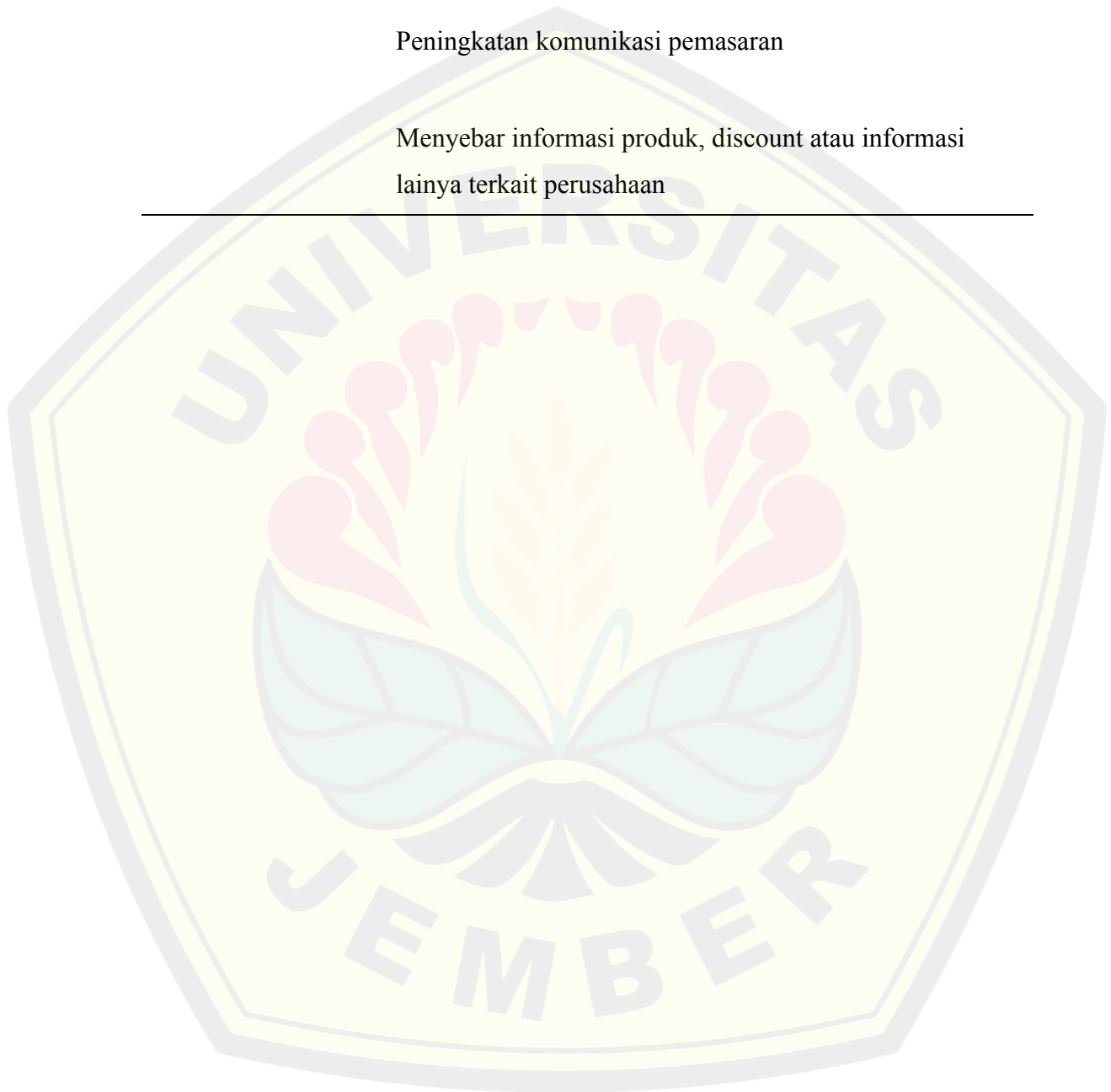
Strategi CRM yang tepat menurut (Parvaneh, Abbasimehr, dan Tarokh 2012) terkait pelanggan dengan kategori potensial customer disarankan perusahaan harus menemukan strategi jarak jauh dengan cara menghubungi melalui telepon, email, fax, dan sebagainya untuk memecahkan masalah menggunakan pengaruh seperti iklan informatif. Iklan informatif memuat beberapaha seperti menginformasikan perubahan harga produk, menginformasikan keberadaan suatu produk, menginformasikan produk atau barang, serta membangun citra perusahaan. Contoh iklan informatif seperti misal toko KPRI UNEJ memiliki produk terbaru yang sedang ramai dan belum banyak menjual, atau Ketika terdapat perubahan harga pihak perusahaan dapat membuat iklan informatif dan disebar kepada pelanggan melalui media sosial, telephone maupun whatsapp.

Strategi diatas dapat disebut sebagai strategi komunikasi. Strategi komunikasi pada hakikatnya merupakan strategi perencanaan dan manajemen komunikasi dalam mencapai suatu tujuan tertentu (Cahyani, Nellyaningsih, dan Marcelino 2022). Komunikasi pemasaran adalah sarana dimana perusahaan menginformasikan, membujuk, dan mengingatkan konsumen secara langsung maupun tidak langsung tentang produk, *discount* atau informasi lainnya terkait perusahaan (Cahyani, Nellyaningsih, dan Marcelino 2022).

Tabel 4.11 Hasil Analisa Cluster 1

<i>Cluster 1</i>	
Nama Kategori	<i>Potensial Customer</i>
<i>Cluster</i>	

Peringkat Cluster	1
Jumlah Anggota	1005
Nama Cluster	<i>Potensial Consumption Customer</i>
Strategi	Peningkatan strategi komunikasi Peningkatan komunikasi pemasaran Menyebarkan informasi produk, discount atau informasi lainnya terkait perusahaan



BAB 5 PENUTUP

5.1 KESIMPULAN

Dari penelitian ini, dapat diperoleh beberapa kesimpulan sebagai berikut:

1. *Cluster* yang terbentuk dari data KPRI UNEJ terdiri dari 2 *cluster* yaitu *cluster* new customer dan *cluster* potensial customer. *cluster* new customer memiliki jumlah anggota pelanggan sebanyak 998 pelanggan. dengan nilai CLV sebesar 0.288768 lebih rendah dari *cluster* potensial customer. *Cluster* potensial customer memiliki jumlah anggota pelanggan sebanyak 1005 pelanggan. dengan nilai CLV sebesar 0.336871 lebih tinggi dari *cluster* new customer
2. *Cluster* new customer perusahaan dapat dengan meningkatkan kepuasan pelanggan, melakukan survei kepuasan pelanggan, evaluasi hasil survei, meningkatkan pelayanan, sedangkan untuk *cluster* potensial customer perusahaan dapat meningkatkan strategi komunikasi, meningkatkan komunikasi pemasaran dan menyebarkan informasi produk, discount atau informasi lainnya terkait perusahaan melalui media sosial ataupun whatsapp.

5.2 SARAN

Selama pengerjaan penelitian ini, tentunya terdapat kekurangan yang dapat diperbaiki dan dikembangkan pada penelitian selanjutnya. Saran yang dapat dipertimbangkan yaitu:

1. Penelitian selanjutnya dapat dikembangkan dengan menggunakan sektor lain dalam proses analisa tidak terbatas terhadap retail.
2. Melakukan penentuan terkait strategi CRM bersama pihak manajer kantor atau perusahaan terkait.

DAFTAR PUSTAKA

- AlizadehZoeram, Ali, dan Ahmadreza KarimiMazidi. 2018. "A New Approach for Customer Clustering by Integrating the LRFM Model and Fuzzy Inference System." *Iranian Journal of Management Studies* 11(2): 351–78.
- Amanah, Siti. 2017. "Peranan Strategi Promosi Pemasaran Terhadap Peningkatan Volume Penjualan." 1: 55–66.
- Bachtiar, Adam Mukharil, Dian Dharmayanti, dan Riky Lutfi Hamzah. 2017. "Penerapan Metode Hierarchical Agglomerative Clustering Untuk Segmentasi Pelanggan Potensial Di Jeger Jersey Indonesia." *Komputa : Jurnal Ilmiah Komputer dan Informatika* 6(1): 35–42.
- Bodet, Guillaume. 2008. "Customer satisfaction and loyalty in service: two concepts, four constructs, several relationships."
- Cahyani, Leni, Nellyaningsih, dan Dandy Marcelino. 2022. 4 วารสารวิชาการมหาวิทยาลัยอีสเทิร์นเอเชีย *How To Improve Understanding Andd Effective Salesmanship Implementation*. Cetakan Pe. Bandung: Widina Bhakti Persada Bandung.
- Chaniago, Harmon, dan Politeknik Negeri Bandung. 2021. 18. *buku-manajemen ritel-harmon-*.
- Gu, Junfei. 2021. "Comparative analysis based on clustering algorithms." *Journal of Physics: Conference Series* 1994(1).
- Gunawan, Indra. 2022. "Bisnis Ritel Kembali Bergairah pada 2022, Ini Buktinya." *Bisnis.com*.
- Gupta, Sunil dkk. 2006. "Modeling customer lifetime value." *Journal of Service Research* 9(2): 139–55.
- Hananto, Valentinus Roby, Agus Dwi Churniawan, dan Ayouvi Poerna Wardhanie. 2017. "Perancangan Analytical CRM untuk Mendukung Segmentasi Pelanggan di Institusi Pendidikan." *Jurnal Ilmiah Teknologi Informasi Asia* 11(1): 79.
- Haryono, Erwin. 2022. "SURVEI PENJUALAN ECERAN APRIL 2022: KINERJA PENJUALAN ECERAN DIPRAKIRAKAN MENINGKAT."

Bank Indonesia.

Hidayati, Rahmatina, Anis Zubair, Aditya Hidayat Pratama, dan Luthfi Indana. 2021. "Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering." *Techno.Com* 20(2): 186–97.

Iriana, Fristiana. 2017. *Metode Penelitian Terapan*. Yogyakarta: Parama Ilmu.

Justitia, Rhayhana Putri, Nurul Hidayat, dan Edy Santoso. 2021. "Implementasi Metode Agglomerative Hierarchical Clustering Pada Segmentasi Pelanggan Barbershop (Studi Kasus: RichDjoe Barbershop Malang)." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* 5(3): 1048–54.

Khotimah, Khusnul. 2015. 1 "EFEKTIVITAS PENGENDALIAN INTERNAL KREDIT PADA KOPERASI PEGAWAI REPUBLIK INDONESIA (KPRI) UNIVERSITAS JEMBER."

Lohonauman, Hans. 2020. "Ipteks Penghitungan Customer Lifetime Value." *Jurnal Ipteks Akuntansi bagi Masyarakat* 4(19–23): 19–23.

Mardiantien, Crisnandra Rahmawati, Imelda Atastina, dan Ibnu Asror. 2020. "Product Segmentation Based on Sales Transaction Data Using Agglomerative Hierarchical Clustering and FMC Model (Case Study: XYZ Company)." *International Conference on Information and Communications Technology (ICOLACT)*.

Marisa, Fitri dkk. 2019. "Segmentation model of customer lifetime value in Small and Medium Enterprise (SMEs) using K-Means Clustering and LRFM model." *International Journal of Integrated Engineering* 11(3): 169–80.

Mohamad, Ismail Bin, dan Dauda Usman. 2013. "Standardization and its effects on K-means clustering algorithm." *Research Journal of Applied Sciences, Engineering and Technology* 6(17): 3299–3303.

Monalisa, Siti. 2018. "Klusterisasi Customer Lifetime Value dengan Model LRFM menggunakan Algoritma K-Means." *Jurnal Teknologi Informasi dan Ilmu Komputer* 5(2): 247.

Mongi, C. E., Y. A.R. Langi, C. E.J.C. Montolalu, dan N. Nainggolan. 2019.

- “Comparison of hierarchical clustering methods (case study: Data on poverty influence in North Sulawesi).” *IOP Conference Series: Materials Science and Engineering* 567(1).
- Müller, Andreas C, dan Sarah Guido. 2017. *Introduction to Machine Learning with Python*. ed. Dawn Schanafelt. United States of America: O’Reilly Media, Inc.
- Nurseptiani, A., Y. Satria, dan H. Burhan. 2021. “Application of agglomerative hierarchical clustering to optimize matching problems in ridesharing for maximize total distance savings.” *Journal of Physics: Conference Series* 1821(1).
- Parvaneh, Amin, Hossein Abbasimehr, dan Mohammad Jafar Tarokh. 2012. “Integrating AHP and Data Mining for Effective Retailer Segmentation Based on Retailer Lifetime Value.” *Journal of Optimization in Industrial Engineering* 11: 25–31.
- Parvaneha, Amin, Hossein Abbasimehrb, dan Mohammad Jafar Tarokhc. 2012. “Integrating AHP and data mining for effective retailer segmentation based on retailer lifetime value.” *Journal of Optimization in Industrial Engineering* 11(September 2012): 25–31.
- Pramesti, Dyang Falila, M. Tanzil Furqon, dan Candra Dewi. 2017. “Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot).” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* 1(9): 723–32.
- Pratikto, Ridzki Okta, dan Natalia Damastuti. 2021. “Klasterisasi Menggunakan Agglomerative Hierarchical Clustering Untuk Memodelkan Wilayah Banjir.” *JOINTECS (Journal of Information Technology and Computer Science)* 6(1): 13.
- Rahadian, Yanuar Rafi, dan Bambang Syairudin. 2020. “Segmentation Analysis of Students in X Course with RFM Model and Clustering.” *Jurnal Sosial Humaniora* 0(1): 59.
- Rahmawati, Nur Fitri, Rangga Dalu Aji Toana, dan Mohammad Sofyan. 2021. “Customer Relationship Marketing Di Ramayana Departement Store Pada Masa Pandemic Covid-19.” *Jurnal Apresiasi Ekonomi* 9(3): 372–78.

- Ramadhani, Lisda, Ika Purnamasari, dan Fidia Deny Tisna Amijaya. 2018. “Penerapan Metode *Complete Linkage* dan Metode Hierarchical Clustering Multiscale Bootstrap (Studi Kasus: Kemiskinan Di Kalimantan Timur Tahun 2016).” *Ekspansional* 9(2016): 1–10. [https://fmipa.unmul.ac.id/files/docs/\[1\] LISDA RAMADHANI 1307015041_Edit.pdf](https://fmipa.unmul.ac.id/files/docs/[1] LISDA RAMADHANI 1307015041_Edit.pdf).
- Reinartz, Werner, Kumar V, dan Mack Robinson J. 2006. *Customer Relationship Management*.
- Suryanto, Edi. 2013. “Implementasi Customer Relationship Management Dengan Market Basket Analysis Pada Toko Buku Online Studi Kasus : Toko Buku Toga Mas.” *Jurnal EKSIS* 6(2): 12–20.
- Widyawati, Widyawati, Wawan Laksito Yuly Saptomo, dan Yustina Retno Wahyu Utami. 2020. “Penerapan Agglomerative Hierarchical Clustering Untuk Segmentasi Pelanggan.” *Jurnal Ilmiah SINUS* 18(1): 75.
- Wijuniamurti, Susi, Sigit Nugroho, dan Ramya Rachmawati. 2022. “Agglomerative Nesting (AGNES) Method and Divisive Analysis (DIANA) Method For Hierarchical Clustering On Some Distance Measurement Concepts.” *Journal of Statistics and Data Science* 1(1): 7–11.
- Zaki, Mohammed J., dan Wagner Jr. Meira. 2020. *Data Mining and Machine Learning Fundamental Concepts and Algorithms*. Second. Cambridge.
- Zoeram, Ali Alizadeh, dan Ahmadreza Karimi Mazidi. 2018. “A New Approach for Customer Clustering by Integrating the LRFM Model and Fuzzy Inference System.” *Iranian Journal of Management Studies* 11(2): 351–78.

LAMPIRAN

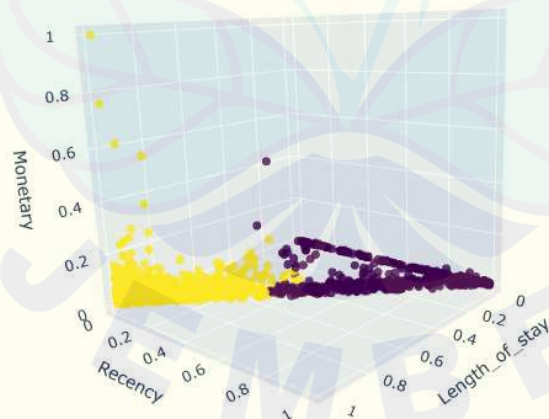
A. Data Hasil Pembentukan Model LRFM



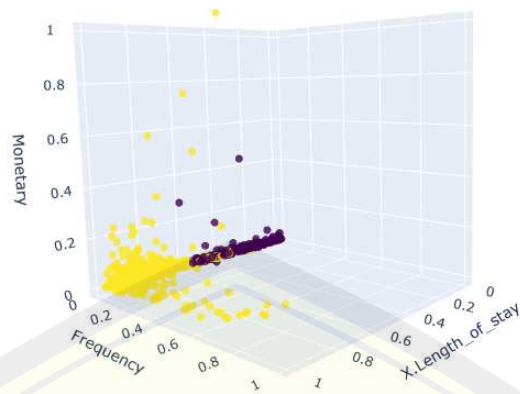
Gambar 6.1 Qr-Code data model LRFM dan hasil cluster

B. Visualisasi Data Dengan Menggunakan 3D

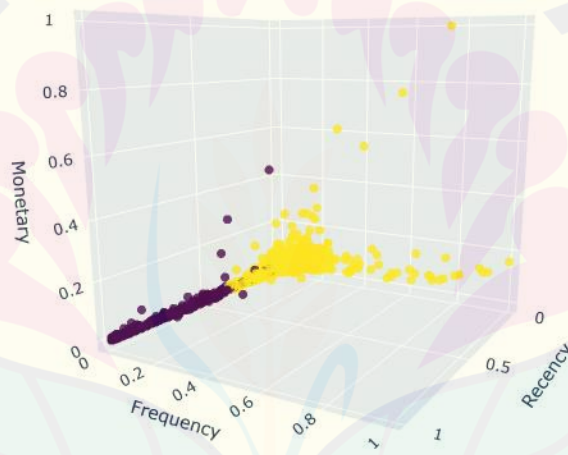
Menampilkan diagram dimensi dengan perbandingan LRM, LFM, RFM



Gambar 6.2 Gambar Hasil Scatter plot 3D model LRM dimana nilai $x = \text{Length_of_stay}$, $y = \text{Recency}$, $z = \text{Monetary}$



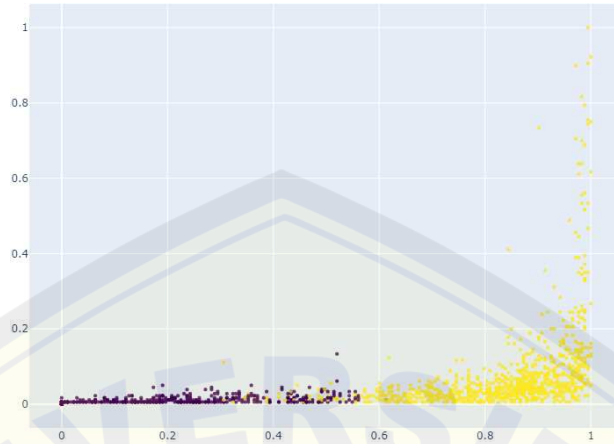
Gambar 6.3 Gambar Hasil Scatter plot 3D model LFM dimana nilai $x = \text{Length_of_stay}$, $y = \text{Frequency}$, $z = \text{Monetary}$



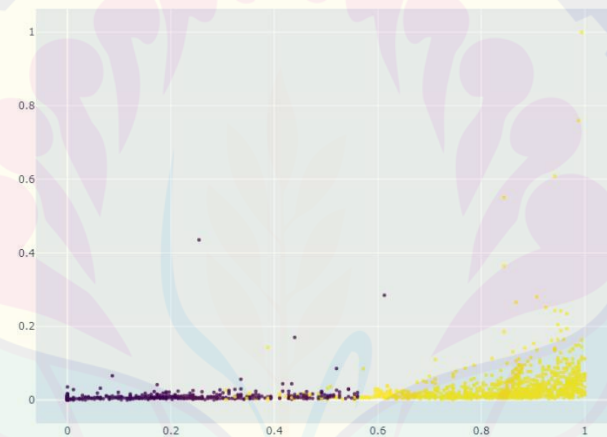
Gambar 6.4 Gambar Hasil Scatter plot 3D model LFM dimana nilai $x = \text{Recency}$, $y = \text{Frequency}$, $z = \text{Monetary}$

C. Visualisasi Data Menggunakan 2D

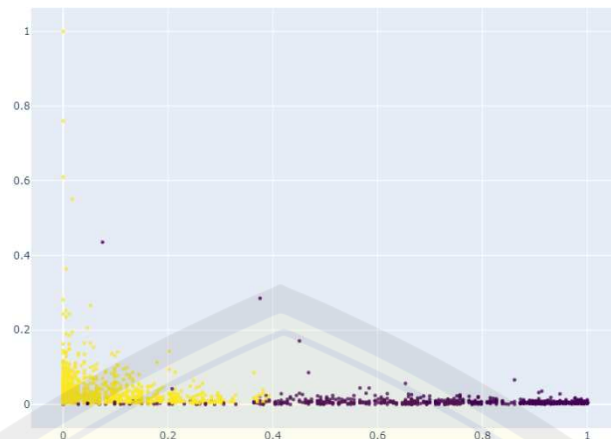
Menampilkan diagram dimensi dengan perbandingan



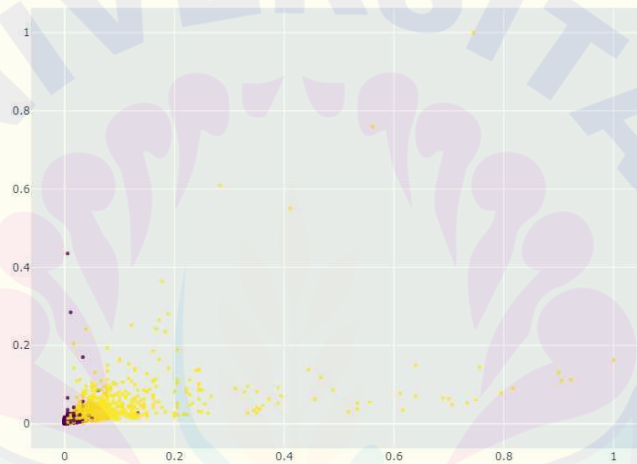
Gambar 6.5 Hasil Scatter plot 2D model LF. dimana nilai $x = \text{Length_of_Stay}$, $y = \text{Frequency}$



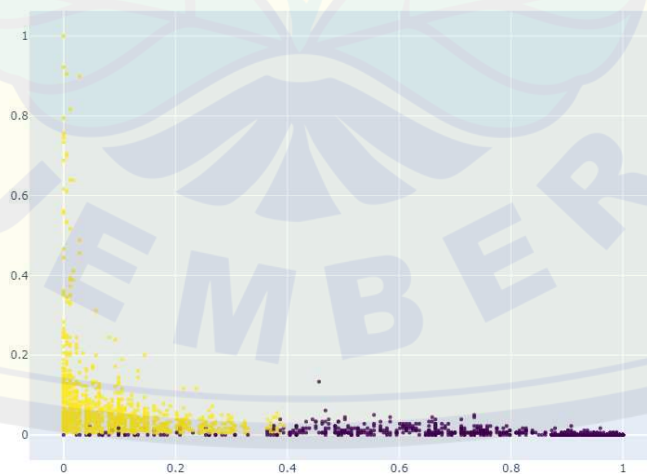
Gambar 6.6 Hasil Scatter plot 2D model LM. dimana nilai $x = \text{Length_of_Stay}$, $y = \text{Monetary}$



Gambar 6.7 Hasil Scatter plot 2D model RM dimana nilai $x = \text{Recency}$, $y = \text{Monetary}$



Gambar 6.8 Hasil Scatter plot 2D model FM dimana nilai $x = \text{Frequency}$, $y = \text{Monetary}$



Gambar 6.9 Hasil Scatter plot 2D model RF dimana nilai $x = \text{Recency}$, $y = \text{Frequency}$

D. Source Code

```
# PEMAHAMAN DATA

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

pip install pandas

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from math import ceil

from sklearn.cluster import KMeans
import plotly.express as px
import plotly.graph_objects as go
from plotly.subplots import make_subplots
from sklearn.metrics import silhouette_samples, silhouette_score
from sklearn.manifold import TSNE
import plotly
plotly.offline.init_notebook_mode(connected=True)

from ipywidgets import interact, interactive, fixed,
interact_manual, VBox, HBox, Layout
import ipywidgets as widgets
import datetime
from datetime import date
import scipy.cluster.hierarchy as sch

dat=pd.read_csv('/kaggle/input/kprifizzz/fixkpri.csv', encoding=
'unicode_escape')
dat.head()

data = dat.copy()

data['InvoiceDate']=pd.to_datetime(data['InvoiceDate'])
data['Sales'] = data.Quantity*data.UnitPrice
data['Year']=data.InvoiceDate.dt.year
data['Month']=data.InvoiceDate.dt.month
data['Week']=data.InvoiceDate.dt.week
#data.InvoiceDate.dt.week
#wk = dt.isocalendar()[1]
data['Year_Month']=data.InvoiceDate.dt.to_period('M')
data['Hour']=data.InvoiceDate.dt.hour
data['Day']=data.InvoiceDate.dt.day
data['is_cancelled']=data.InvoiceNo.apply(lambda x: 'Yes' if
x[0]=='C' else 'No')
data['weekday'] = data.InvoiceDate.dt.day_name()
data['Quarter'] = data.Month.apply(lambda m:'Q'+str(ceil(m/4)))
data['Date']=pd.to_datetime(data[['Year', 'Month', 'Day']])
data.head(150)
```

```

df = px.data.gapminder()
df = df [['country', 'iso_alpha']]
data =
pd.merge(data, df[['country', 'iso_alpha']], left_on='Country', right_
on='country', how='left').drop(columns=['country'])
print(data)
# del df

data_=data[data.is_cancelled=='No']
del data

sales_by_date =
data_.groupby(by='Date')['Sales'].sum().reset_index()
fig =
go.Figure(data=go.Scatter(x=sales_by_date.Date, y=sales_by_date.Sal
es
, line = dict(color='black',
width=1.5)))
fig.update_layout(xaxis_title="Date", yaxis_title="Sales", title='Da
ily Sales', template='ggplot2')
fig.show()

customer_by_month1 =
data_.groupby('CustomerId')['Date'].min().reset_index()
customer_by_month1['days'] =
pd.TimedeltaIndex(customer_by_month1.Date.dt.day, unit="D")
customer_by_month1['Month'] = customer_by_month1.Date-
customer_by_month1.days+pd.DateOffset(days=1)
customer_by_month1['Quarter_acquisition'] =
customer_by_month1['Month'].dt.quarter.apply(lambda x:'Q'+str(x))
customer_by_month1['Year_acquisition'] =
customer_by_month1['Month'].dt.year
customer_by_month = data_.groupby(by =
customer_by_month1.Month)['CustomerId'].size().reset_index()
customer_by_month.sort_values(by
='Month', ascending=True, inplace=True)
customer_by_month['cum_customer'] =
np.cumsum(customer_by_month.CustomerId)
customer_by_month['Month_1'] =
customer_by_month['Month'].dt.strftime('%b-%y')

plt.style.use('ggplot')
plt.figure(figsize=(20,5))
plt.plot(customer_by_month.Month_1, customer_by_month.cum_customer,
'bo-', color='black')

for d,c in
zip(customer_by_month['Month_1'], customer_by_month['cum_customer']
):

    label = "{:.0f}".format(c)

    plt.annotate(label,
                 (d,c),
                 textcoords="offset points"

```

```

        , bbox=dict(boxstyle="round", fc="none",
ec="gray")
        #,arrowprops=dict(arrowstyle="-
",connectionstyle="angle,angleA=0,angleB=10,rad=90")
        ,xytext=(0,10),
        ha='center')
plt.show()

del customer_by_month

sales_by_hour =
data_.groupby(by='Hour')['Sales'].sum().reset_index()
sales_by_weekday =
data_.groupby(by='weekday')['Sales'].sum().reset_index()

fig = make_subplots(rows=1, cols=2, subplot_titles=("Total Hourly
Sales", "Total Sales by Weekday"))
fig.add_trace(go.Bar(y=sales_by_hour.Hour,
x=sales_by_hour.Sales,orientation='h'),row=1, col=1)
fig.add_trace(go.Bar(x=sales_by_weekday.weekday,
y=sales_by_weekday.Sales),row=1, col=2)
fig.update_layout(height=700, width=800,template='ggplot2')
fig.update_xaxes(title_text="Sales", row=1, col=1)
fig.update_xaxes(title_text="Weekday", row=1, col=2)
fig.update_yaxes(title_text="Hours", row=1, col=1)
fig.update_yaxes(title_text="Sales", row=1, col=2)
fig.show()

del [sales_by_hour,sales_by_weekday]

LRFM =
data_.groupby('CustomerId').agg(Frequency=pd.NamedAgg(column="Invo
iceNo", aggfunc="nunique")
                                ,Monetary=pd.NamedAgg(colu
mn="Sales", aggfunc="sum"),
                                Recency =
pd.NamedAgg(column='InvoiceDate',aggfunc='min')).reset_index()
length = data_.groupby('CustomerId')['Date'].max() -
data_.groupby('CustomerId')['Date'].min()
length = (length/np.timedelta64(1, 'D')).reset_index()
length.columns = ['CustomerId','Length_of_stay']

LRFM = LRFM.merge(length,on='CustomerId',how='inner')
del length

LRFM.head(100)

e = LRFM['Recency'].min()
print('minimun :'+ str(e))

LRFM['Recency'] = LRFM['Recency'].apply(lambda x : (x - e).days)

LRFM.head(10)

a=(LRFM['Frequency'].max())

```

```

b=(LRFM['Frequency'].min())
print(a,b)

c= (LRFM['Monetary'].max())
d= (LRFM['Monetary'].min())
print(c,d)

e= (LRFM['Recency'].max())
f= (LRFM['Recency'].min())
print(e,f)

g= (LRFM['Length_of_stay'].max())
h= (LRFM['Length_of_stay'].min())
print(g,h)

LRFM.head(15)

#urutan
LRFM_1= LRFM[['Length_of_stay',"Recency","Frequency","Monetary"]]
LRFM_1.head()

from sklearn import preprocessing

min_max_scaler = preprocessing.MinMaxScaler(feature_range=(0,1))
#inisialisasi normalisasi MinMax
LRFM_2 = min_max_scaler.fit_transform(LRFM_1) #transformasi MinMax
untuk fitur

print("dataset setelah dinormalisasi :")
print(LRFM_2)

LRFM_3= pd.DataFrame(LRFM_2)

print(LRFM_3)

df = pd.DataFrame(dat)
CustomerId = []
#df.insert(0, column="CustomerId", value=range(2002))
LRFM_3.columns =
["Length_of_stay","Recency","Frequency","Monetary"]

LRFM_3.head()

LRFM_3.to_csv('LRFM_3.csv',float_format='%.2f')

plt.figure(figsize=(10, 7))
plt.title("Dendrograms")
dend = sch.dendrogram(sch.linkage(LRFM_3, method='ward'))

from sklearn.cluster import AgglomerativeClustering
hc = AgglomerativeClustering(n_clusters = 10, affinity =
'euclidean', linkage = 'ward')
y_hc = hc.fit_predict(LRFM_3)

# Silhouette analysis

```

```

range_n_clusters = [2,3,4,6]

for num_clusters in range_n_clusters:

    hc = AgglomerativeClustering(n_clusters = num_clusters ,
affinity = 'euclidean', linkage = 'ward') #inisialisasi hc
digunakan untuk mencari cluster berdasarkan num_clusters
    y_hc = hc.fit_predict(LRFM_3)
    cluster_labels = hc.labels_

    # silhouette score
    silhouette_avg = silhouette_score(LRFM_3, cluster_labels)
    print("For n_clusters={0}, the silhouette score is
{1}".format(num_clusters, silhouette_avg))

X=LRFM_3
cluster_lbls = AgglomerativeClustering(n_clusters=2,
affinity='euclidean', linkage='ward').fit_predict(X)
X['cluster'] = cluster_lbls
X['sample_silhouette_values'] = silhouette_samples(X,
cluster_lbls)
X['txt']=X.cluster.apply(lambda x:'Cluster '+str(x))

df = X.groupby('cluster').agg({'cluster':'size',
'Monetary':'mean', 'Frequency':'mean', 'Recency':'mean', 'Length_of_s
tay':'mean'}) \
    .rename(columns={'cluster':'Size', 'Monetary':'Avg
Sales', 'Frequency':'Avg Frequency', 'Recency':'Avg
Recency', 'Length_of_stay':'Avg Lenght of Stay'}) \
    .reset_index().sort_values(by = 'Avg Sales')

cluster_map={'Cluster 0':'yellow', 'Cluster 1':'purple'}

txt =['Size = {0:.0f}'.format(i) for i in df.Size]
df['cluster']=df.cluster.apply(lambda x:'Cluster '+str(x))
df['Group']=df.cluster.map(cluster_map)

fig = make_subplots(rows=1, cols=4, subplot_titles=("Avg Sales",
"Avg Frequency", 'Avg Recency', 'Avg Lenght of Stay'))

fig.add_trace(go.Bar(y=df.cluster, x=df['Avg Sales'], hovertext=txt
, text=txt, textposition='auto', marker_color
=df.Group, orientation='h'), row=1, col=1)
fig.add_trace(go.Bar(y=df.cluster, x=df['Avg
Frequency'], hovertext=txt
, text=txt, textposition='auto', marker_color
=df.Group, orientation='h'), row=1, col=2)
fig.add_trace(go.Bar(y=df.cluster, x=df['Avg
Recency'], hovertext=txt
, text=txt, textposition='auto', marker_color
=df.Group, orientation='h'), row=1, col=3)
fig.add_trace(go.Bar(y=df.cluster, x=df['Avg Lenght of
Stay'], hovertext=txt
, text=txt, textposition='auto', marker_color
=df.Group, orientation='h'), row=1, col=4)

```

```

fig.update_traces(marker_line_color='rgb(8,48,107)',
marker_line_width=1.5, opacity=0.8)
fig.update_layout(title_text='Cluster Size',width =
800,height=600,template='ggplot2'
,font=dict(family="Courier New,
monospace",size=10,color="RebeccaPurple"))

fig.show()

#figure LRF
#3 D view of clusters
fig =
go.Figure(data=[go.Scatter3d(x=X.Length_of_stay,y=X.Recency,z=X.Fr
equency,mode='markers'
,marker=dict(size=4,color=X.clu
ster
,colorscale='Virid
is',opacity=0.8))])

# tight layout
fig.update_layout(margin=dict(l=1, r=2, b=1, t=1)
,scene=dict(xaxis=dict(title='Length_of_stay')
,yaxis=dict(title='Recency')
,zaxis=dict(title='Frequency')),widt
h=700,height=500)
fig.show()

#figure LRM
#3 D view of clusters
fig =
go.Figure(data=[go.Scatter3d(x=X.Length_of_stay,y=X.Recency,z=X.Mo
netary,mode='markers'
,marker=dict(size=4,color=X.clu
ster
,colorscale='Virid
is',opacity=0.8))])

# tight layout
fig.update_layout(margin=dict(l=1, r=2, b=1, t=1)
,scene=dict(xaxis=dict(title='Length_of_stay')
,yaxis=dict(title='Recency')
,zaxis=dict(title='Monetary')),width
=700,height=500)
fig.show()

#figure LFM
#3 D view of clusters
fig =
go.Figure(data=[go.Scatter3d(x=X.Length_of_stay,y=X.Frequency,z=X.
Monetary,mode='markers'
,marker=dict(size=4,color=X.clu
ster
,colorscale='Virid
is',opacity=0.8))])

# tight layout

```



```

fig.update_layout(margin=dict(l=1, r=2, b=1, t=1)
                  ,scene=dict(xaxis=dict(title='X.Length_of_stay')
                              ,yaxis=dict(title='Frequency')
                              ,zaxis=dict(title='Monetary')),width
=700,height=500)
fig.show()

#figure RFM
#3 D view of clusters
fig =
go.Figure(data=[go.Scatter3d(x=X.Recency,y=X.Frequency,z=X.Monetary,mode='markers'
                              ,marker=dict(size=4,color=X.clusters
                              ,colorscale='Viridis',opacity=0.8))])

# tight layout
fig.update_layout(margin=dict(l=1, r=2, b=1, t=1)
                  ,scene=dict(xaxis=dict(title='Recency')
                              ,yaxis=dict(title='Frequency')
                              ,zaxis=dict(title='Monetary')),width
=700,height=500)
fig.show()

#figure LR
#2 D view of clusters
fig =
go.Figure(data=[go.Scatter(x=X.Length_of_stay,y=X.Recency,mode='markers'
                              ,marker=dict(size=4,color=X.clusters
                              ,colorscale='Viridis',opacity=0.8))])

# tight layout
fig.update_layout(margin=dict(l=1, r=2, b=1, t=1)
                  ,scene=dict(xaxis=dict(title='Length_of_stay')
                              ,yaxis=dict(title='Recency')),width=
700,height=500)
fig.show()

#figure LF
#2 D view of clusters
fig =
go.Figure(data=[go.Scatter(x=X.Length_of_stay,y=X.Frequency,mode='markers'
                              ,marker=dict(size=4,color=X.clusters
                              ,colorscale='Viridis',opacity=0.8))])

# tight layout
fig.update_layout(margin=dict(l=1, r=2, b=1, t=1)
                  ,scene=dict(xaxis=dict(title='Length_of_stay')

```

```

, yaxis=dict(title='Frequency')), width
h=700,height=500)
fig.show()

#figure LM
#2 D view of clusters
fig =
go.Figure(data=[go.Scatter(x=X.Length_of_stay,y=X.Monetary,mode='m
arkers'
ster
,marker=dict(size=4,color=X.clu
ster
,colorscale='Virid
is',opacity=0.8))])

# tight layout
fig.update_layout(margin=dict(l=1, r=2, b=1, t=1)
,scene=dict(xaxis=dict(title='Lenght_of_Stay')
,yaxis=dict(title='Monetary')),width
=700,height=500)
fig.show()

#figure RM
#2 D view of clusters
fig =
go.Figure(data=[go.Scatter(x=X.Recency,y=X.Monetary,mode='markers'
ster
,marker=dict(size=4,color=X.clu
ster
,colorscale='Virid
is',opacity=0.8))])

# tight layout
fig.update_layout(margin=dict(l=1, r=2, b=1, t=1)
,scene=dict(xaxis=dict(title='Recency')
,yaxis=dict(title='Monetary')),width
=700,height=500)
fig.show()

#figure RM
#2 D view of clusters
fig =
go.Figure(data=[go.Scatter(x=X.Frequency,y=X.Monetary,mode='marker
s'
ster
,marker=dict(size=4,color=X.clu
ster
,colorscale='Virid
is',opacity=0.8))])

# tight layout
fig.update_layout(margin=dict(l=1, r=2, b=1, t=1)
,scene=dict(xaxis=dict(title='Frequency')
,yaxis=dict(title='Monetary')),width
=700,height=500)
fig.show()

#figure RF
#2 D view of clusters

```

```

fig =
go.Figure(data=[go.Scatter(x=X.Recency,y=X.Frequency,mode='markers
',
                                ,marker=dict(size=4,color=X.clu
ster
                                ,colorscale='Virid
is',opacity=0.8))])

# tight layout
fig.update_layout(margin=dict(l=1, r=2, b=1, t=1)
                  ,scene=dict(xaxis=dict(title='Recency')
                              ,yaxis=dict(title='Frequency'))),width
h=700,height=500)
fig.show()

Tabel = pd.DataFrame (df)

print(Tabel)

#fddd =pd.DataFrame({'Avg Sales', 'Avg Recency', 'Avg Recency',
'Avg Lenght of Stay')
sk = Tabel[['Avg Sales', 'Avg Frequency', 'Avg Recency', 'Avg
Lenght of Stay']]

sk.values

# PERKALIAN CLV
#PR OBJECT TO ARRAY
blok1 = sk.values
blok2 = [0.238,0.088,0.326,0.348] #urutannya dibalik
nilaiclv1 = []
print (blok1.dtype)

for x in blok1:
    nilaiclv =[]
    for z,i in enumerate(x):
        nilaiclv.append(i*blok2[z])
    nilaiclv1.append(nilaiclv)
print (nilaiclv1)

nilaiclvdeal = pd.DataFrame(nilaiclv1)

print(nilaiclvdeal)

```