

# SKRIPSI

## ANALISIS SEGMENTASI PELANGGAN MENGGUNAKAN ALGORITMA *FUZZY C-MEANS* DAN MODEL *RECENCY FREQUENCY MONETARY*

Oleh :

Yusuf Fadilah Rukmana

0651 18 220



**PROGRAM STUDI ILMU KOMPUTER  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS PAKUAN  
BOGOR  
2023**

## HALAMAN PERSEMBAHAN

*“Strength is the product of struggle. You must do what others don't to achieve what others won't.”*

*- Henry Rollins –*

*Ucapan terima kasih ini saya persembahkan kepada semua individu yang telah memberikan dukungan, bantuan, dan inspirasi selama proses penyelesaian skripsi ini.*

*Pertama, terima kasih kepada orang tua dan keluarga saya atas dukungan, cinta, dan motivasi tanpa batas. Kalian selalu menjadi pilar dalam hidup saya dan memberikan dukungan tak tergoyahkan.*

*Terima kasih kepada dosen pembimbing saya atas bimbingan, dorongan, dan wawasan yang berharga. Terima kasih atas kesabaran dan dedikasinya dalam membimbing langkah-langkah penelitian ini.*

*Saya juga ingin mengucapkan terima kasih kepada teman-teman dan rekan-rekan sejawat yang telah memberikan dukungan moral dan sharing pengetahuan selama perjalanan ini. Kalian memberikan inspirasi dan keceriaan yang membuat proses ini lebih berarti.*

*Saya percaya bahwa kecerdasan tidak dapat diukur semata dari sebuah karya tulis. Kecerdasan melibatkan lebih dari itu. Ia melibatkan kreativitas, kemampuan berpikir kritis, ketekunan, dan ketangguhan dalam menghadapi tantangan. Skripsi ini adalah cermin dari upaya yang saya lakukan untuk menggali pengetahuan dan mengembangkan pemahaman saya dalam bidang ini. Ia mewakili kerja keras saya, kerjasama dengan para pembimbing, dan dukungan dari orang-orang terdekat yang selalu ada untuk saya.*

## HALAMAN PENGESAHAN

Judul : Analisis Segmentasi Pelanggan Menggunakan Algoritma *Fuzzy C-Means*  
dan *Model Recency, Frequency, Monetary*.

Nama : Yusuf Fadilah Rukmana

Npm : 065118220

Mengesahkan,

Pembimbing Pendamping

FMIPA - UNPAK

Irma Anggraeni, M.Kom.

Pembimbing Utama

FMIPA - UNPAK

Asep Denih, S.Kom., M.Sc., Ph.D.

Mengetahui,

ζetua Program Studi Ilmu Komputer

FMIPA - UNPAK

Arie Qur'ania, M.Kom.

Dekan

FMIPA - UNPAK

Asep Denih, S.Kom., M.Sc., Ph.D.

## **PERNYATAAN KEASLIAN KARYA TULIS SKRIPSI**

Dengan ini saya Yusuf Fadilah Rukmana, menyatakan bahwa:

Nama : Yusuf Fadilah Rukmana

Npm : 065118220

Program Studi : Ilmu Komputer / S1

Fakultas : Matematika dan Ilmu Pengetahuan Alam Universitas Pakuan Bogor

Judul Skripsi : “Analisis Segmentasi Pelanggan Menggunakan *Algoritma Fuzzy C-Means* dan Model *Recency, Frequency, Monetary*”.

Saya menyatakan dengan sebenar-benarnya bahwa karya tulis ini tidak pernah dipublikasikan atau digunakan untuk mendapatkan gelar sarjana di universitas lain, kecuali pada bagian-bagian tertentu yang sumber informasinya telah dicantumkan dengan cara referensi yang semestinya. Apabila terdapat gugatan di kemudian hari, saya bersedia dikenakan sanksi sesuai dengan peraturan yang berlaku.

Bogor, 21 Juni 2023

Yusuf Fadilah Rukmana

## **PERNYATAAN PELIMPAHAN SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA**

---

Saya, yang bertanda tangan di bawah ini:

Nama: Yusuf Fadilah Rukmana

NPM: 065118220

Dengan ini menyatakan bahwa skripsi yang berjudul "Analisis Segmentasi Pelanggan menggunakan Algoritma *Fuzzy C-Means* dan model *Recency Frequency Monetary*" adalah hasil karya saya sendiri dengan bimbingan dari komisi pembimbing. Skripsi ini belum pernah diajukan dalam bentuk apa pun kepada perguruan tinggi manapun.

Sumber informasi yang saya kutip atau gunakan dari karya yang telah diterbitkan maupun yang belum diterbitkan dari penulis lain telah diakui dan dicantumkan dalam teks dan daftar pustaka skripsi ini. Saya memberikan hak paten dan hak cipta dari karya tulis ini kepada Universitas Pakuan.

Demikian pernyataan ini saya buat dengan sebenarnya, sebagai bentuk kejujuran dan integritas terhadap karya ilmiah yang saya hasilkan.

Bogor, 21 Juni 2023

Yusuf Fadilah Rukmana

## RIWAYAT HIDUP



**Yusuf Fadilah Rukmana**, lahir di Bogor pada tanggal 18 November 2000 sebagai anak kedua dari pasangan Bapak Nana Rukmana dan Ibu Iis Purnawati. Saya memulai pendidikan di Sekolah Dasar Negeri Kartika Sejahtera 01 pada tahun 2006 dan berhasil menyelesaikan pendidikan tersebut pada tahun 2012. Selanjutnya, saya melanjutkan pendidikan ke jenjang Sekolah Menengah Pertama di SMP Negeri 1 Tajurhalang dari tahun 2012 hingga 2015. Pada tahun 2015, saya melanjutkan ke jenjang pendidikan yang lebih tinggi, yakni Sekolah Menengah Atas di SMAN 1 Tajurhalang, dan berhasil lulus pada tahun 2018.

Memulai pendidikan di Perguruan Tinggi pada tahun 2018 untuk melanjutkan ke jenjang pendidikan yang lebih tinggi merupakan langkah yang saya ambil. Karena minat saya dalam bidang informatika, saya memilih Program Studi Ilmu Komputer di salah satu universitas di Bogor. Biografi singkat ini saya susun untuk memenuhi persyaratan kelulusan penelitian Tugas Akhir/Skripsi.

Saya melakukan Penelitian Tugas Akhir/Skripsi di Toko Tas Hallogen Bogor dengan judul "Analisis Segmentasi Pelanggan Menggunakan Algoritma *Fuzzy C-Means* dan Model *Recency Frequency Monetary*". Penelitian ini memakan waktu beberapa bulan dan dalam penelitian tersebut, saya mengumpulkan dan menganalisis data dari toko tas tersebut, serta melakukan penelitian menggunakan algoritma *Fuzzy C-Means* dan model *Recency Frequency Monetary* untuk menganalisis segmentasi pelanggan.

## RINGKASAN

Toko Hallogen Bogor merupakan usaha aksesoris yang telah memanfaatkan teknologi untuk meningkatkan proses bisnisnya, dengan memanfaatkan media sosial dan aplikasi *e-commerce*. Toko ini fokus pada tiga jenis tas, yaitu *Handbag*, *Sling bag*, dan Tas Ransel dengan sistem penjualan *Ready Stock*. Awalnya, Hallogen Bogor berfokus pada produk Tas Ransel, namun kemudian mengembangkan bisnisnya dengan menambahkan produk aksesoris lainnya untuk menarik minat pelanggan. Pelanggan memiliki peran yang sangat penting bagi setiap perusahaan. Mereka bukan hanya sumber keuntungan, namun juga aset yang harus dikelola dengan baik. Oleh karena itu, segmentasi pelanggan menjadi hal yang penting untuk memahami perilaku konsumen dan menerapkan strategi pemasaran yang tepat untuk meningkatkan pendapatan perusahaan. *Data mining* menjadi solusi untuk melakukan segmentasi konsumen, dengan mengelola data yang tersimpan dalam suatu basis data menjadi pengetahuan baru yang berharga. Mengelompokkan pelanggan dengan kesamaan karakteristik dapat membantu dalam pengambilan keputusan yang tepat, karena setiap pelanggan memiliki kebutuhan, preferensi, harapan, dan perilaku yang berbeda-beda. Perlakuan yang sama terhadap semua pelanggan dapat berdampak negatif bagi perusahaan, karena pelanggan yang tidak begitu bernilai bagi perusahaan bisa menjadi penghancur nilai daripada pencipta nilai bagi perusahaan. Dengan segmentasi pelanggan yang tepat, perusahaan dapat mengoptimalkan pemasaran dan penjualan produknya. *Data mining* juga dapat membantu perusahaan untuk mengekstrak pengetahuan yang berharga dari data yang tersedia, sehingga dapat meningkatkan efektivitas bisnis dan mengambil keputusan yang lebih cerdas.

## KATA PENGANTAR

Puji syukur kami panjatkan kepada Tuhan Yang Maha Esa, karena atas rahmat dan karunia-Nya saya telah menyelesaikan laporan penelitian yang berjudul "Analisis Segmentasi Pelanggan Menggunakan Algoritma *Fuzzy C-Means* dan Model *Recency, Frequency, Monetary*" sesuai dengan yang direncanakan dan dapat terselesaikan dengan baik. Tak lupa, sholawat beserta salam kita haturkan kepada Nabi Besar Muhammad SAW, sebagaimana kita yang menjadi pengikutnya hingga akhir zaman.

Skripsi ini diajukan untuk memenuhi syarat kelulusan mata kuliah Skripsi di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Pakuan Bogor. Tidak dapat disangkal bahwa diperlukan usaha yang keras dalam penyelesaian pengerjaan skripsi ini. Saya menyadari sepenuhnya bahwa tanpa adanya bantuan dari berbagai pihak, saya akan banyak menemui kesulitan dalam penyusunan laporan ini. Maka dari itu, tak lupa ucapan terima kasih yang sebesar-besarnya kepada:

1. Asep Denih, S.Kom., M.Sc., Ph.D., selaku Dosen Pembimbing Utama yang memberikan bimbingan, saran, dan ide yang sangat berperan penting dalam menyelesaikan penelitian ini.
2. Irma Anggraeni, M.Kom., selaku Dosen Pembimbing Pendamping yang juga berperan penting dalam memberikan dukungan serta masukan dalam proses penelitian.
3. Arie Qurania, M.Kom., selaku Ketua Program Studi Ilmu Komputer Fakultas Matematika Ilmu Pengetahuan Alam.
4. Kedua orangtua dan keluarga yang tiada henti memberikan semangat serta doa selama proses penelitian ini berlangsung.
5. Seluruh teman seperjuangan yang sudah ikut andil dan membantu dalam proses penelitian, saya ucapkan terima kasih yang sebesar-besarnya. Tanpa bantuan dan dukungan kalian, saya tidak akan berhasil menyelesaikan penelitian ini dengan baik.

Saya menyadari bahwa laporan ini belum sempurna. Oleh karena itu, saran dan kritik yang membangun dari rekan-rekan sangat dibutuhkan untuk penyempurnaan laporan ini. Akhir kata, semoga laporan ini dapat bermanfaat bagi pembaca dan semoga Allah SWT membalas semua kebaikan dan selalu mencurahkan taufik-Nya, amin ya Robbal 'alamin.

Bogor, 21 Juni 2023

Yusuf Fadilah Rukmana

## DAFTAR ISI

HALAMAN PERSEMBAHAN .....	i
HALAMAN PENGESAHAN .....	ii
PERNYATAAN KEASLIAN KARYA TULIS SKRIPSI.....	iii
PERNYATAAN PELIMPAHAN SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA .....	iv
RIWAYAT HIDUP .....	v
RINGKASAN.....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI.....	vii
DAFTAR GAMBAR.....	x
DAFTAR TABEL.....	xi
DAFTAR LAMPIRAN.....	xii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang .....	1
1.2 Tujuan .....	2
1.3 Ruang Lingkup.....	2
1.4 Manfaat Penelitian .....	2
BAB II TINJAUAN PUSTAKA .....	3
2.1 Tinjauan Pustaka .....	3
2.1.1 Pelanggan.....	3
2.1.2 Segmentasi Pelanggan .....	3
2.1.3 Model Recency, Frequency, Monetary.....	3
2.1.4 Clustering.....	4
2.1.5 Algoritma Fuzzy C-Means.....	4
2.1.6 Elbow Method.....	5
2.1.7 Davies-Bouldin Index (DBI) .....	6
2.2 Penelitian Terdahulu .....	7
2.3 Tabel Perbandingan.....	8
BAB III METODOLOGI PENELITIAN .....	9
3.1 Metode Penelitian.....	9
3.1.1 Selection.....	9

3.1.2	Pre-processing.....	9
3.1.3	Transformation.....	9
3.1.4	Mining.....	9
3.1.5	Interpretation / Evaluation .....	9
3.1.6	Knowledge .....	10
3.2	Waktu dan Tempat Penelitian .....	10
3.3	Alat dan Bahan .....	10
3.3.1	Alat Penelitian.....	10
3.3.2	Bahan Penelitian .....	10
<b>BAB IV PERANCANGAN DAN IMPLEMENTASI .....</b>		<b>11</b>
4.1	Perancangan .....	11
4.2	Implementasi .....	11
4.2.1	Data Selection .....	11
4.2.2	Data Preprocessing .....	12
4.2.3	Data Transformation .....	12
4.2.4	Metode Fuzzy C-Means.....	16
4.2.5	Evaluasi.....	21
4.2.6	Knowledge .....	23
4.3	Tahapan Implementasi Menggunakan Visual Studio Code.....	23
<b>BAB V HASIL DAN PEMBAHASAN .....</b>		<b>24</b>
5.1	Hasil .....	24
5.1.1	Tahap Pengumpulan Data .....	24
5.1.2	Hasil Data Selection.....	24
5.1.3	Hasil Data Preprocessing .....	24
5.1.4	Hasil Data Transformasi .....	25
5.1.5	Hasil Analisis Segmentasi Pelanggan.....	25
5.2	Pengukuran Keakuratan Hasil Cluster .....	27
5.3	Pembahasan.....	29
<b>BAB VI KESIMPULAN DAN SARAN .....</b>		<b>31</b>
6.1	Kesimpulan .....	31
6.2	Saran.....	31
<b>DAFTAR PUSTAKA .....</b>		<b>32</b>
<b>LAMPIRAN.....</b>		<b>34</b>

## DAFTAR GAMBAR

<b>Gambar 1.</b> Knowledge Discovery in Database (KDD) .....	9
<b>Gambar 2.</b> Data Pelanggan yang digunakan .....	11
<b>Gambar 3.</b> Algoritma <i>Fuzzy C-Means</i> .....	17
<b>Gambar 4.</b> Evaluasi Cluster Elbow Method.....	22
<b>Gambar 5.</b> Implementasi <i>Visual Studio Code</i> pada Analisis Data .....	23
<b>Gambar 6.</b> Implementasi <i>Visual Studio Code</i> pada Proses Pengolahan Data .....	23
<b>Gambar 7.</b> RFM Model Score .....	26
<b>Gambar 8.</b> Hasil Perhitungan dengan Algoritma <i>Fuzzy C-Means</i> .....	27
<b>Gambar 9.</b> Penempatan Label <i>Cluster</i> pada Data RFM Normalisasi.....	27
<b>Gambar 10.</b> Hasil Pengelompokan Data dalam Kelompok-Kelompok ( <i>Cluster</i> )....	29
<b>Gambar 11.</b> Jumlah Data dalam Setiap Kelompok <i>Cluster</i> .....	30
<b>Gambar 12.</b> Hasil Pengelompokan Data dalam Bentuk Plot (Grafik) .....	30

## DAFTAR TABEL

<b>Tabel 1.</b> Tabel Perbandingan.....	8
<b>Tabel 2.</b> <i>Data Cleaning</i> .....	12
<b>Tabel 3.</b> <i>Recency</i> .....	13
<b>Tabel 4.</b> <i>Frequency</i> .....	14
<b>Tabel 5.</b> <i>Monetary</i> .....	14
<b>Tabel 6.</b> <i>RFM Score</i> .....	15
<b>Tabel 7.</b> Normalisasi <i>RFM Score</i> .....	16
<b>Tabel 8.</b> Tabel Penentuan <i>Cluster</i> .....	17
<b>Tabel 9.</b> Inisialisasi <i>Random Awal</i> .....	18
<b>Tabel 10.</b> Pusat <i>Cluster Awal</i> .....	19
<b>Tabel 11.</b> Fungsi Objektif Awal.....	19
<b>Tabel 12.</b> Iterasi Awal .....	20
<b>Tabel 13.</b> Iterasi Akhir .....	20
<b>Tabel 14.</b> Tabel <i>Cluster</i> .....	21
<b>Tabel 15.</b> Hasil 2 <i>Cluster</i> .....	22
<b>Tabel 16.</b> Tabel Model RFM.....	26
<b>Tabel 17.</b> Nilai SSW .....	28
<b>Tabel 18.</b> Nilai SSB.....	28

## DAFTAR LAMPIRAN

<b>Lampiran 1.</b> Surat Keterangan Penelitian .....	35
<b>Lampiran 2.</b> Surat Perizinan Perusahaan .....	37
<b>Lampiran 3.</b> Data Penjualan Toko Hallogen.....	39
<b>Lampiran 4.</b> Visualisasi Website .....	40

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Toko Hallogen Bogor adalah sebuah usaha yang bergerak di bidang aksesoris, yang menyediakan berbagai produk seperti tas, gantungan kunci, dan dompet. Untuk meningkatkan proses bisnisnya, toko ini telah memanfaatkan teknologi dengan menggunakan media sosial serta aplikasi *e-commerce* yaitu aplikasi Shopee. Toko Hallogen Bogor memfokuskan penjualan pada tiga jenis tas, yaitu *Handbag*, *Sling bag*, dan Tas Ransel. Sistem penjualan yang digunakan adalah sistem *Ready Stock*, yang artinya setiap produk selalu tersedia untuk dijual. Toko Hallogen Bogor didirikan pada tahun 2018 dengan fokus awal pada produk Tas Ransel. Namun, setelah beberapa tahun beroperasi, Hallogen Bogor mengembangkan bisnisnya dengan menambahkan beberapa produk aksesoris lainnya agar lebih menarik bagi pelanggan. Meskipun begitu, Hallogen Bogor masih mengalami kesulitan dalam menentukan pelanggan potensial, padahal pelanggan memiliki posisi yang sangat penting bagi setiap perusahaan.

Pelanggan merupakan sumber keuntungan dan kehidupan bagi perusahaan seperti halnya pada bisnis di bidang ritel. Pelanggan merupakan aset bagi perusahaan. Saat ini, perusahaan berlomba-lomba dalam memenangkan hati para pelanggan. Karena masing-masing pelanggan memiliki kebutuhan, preferensi, harapan, dan perilaku yang berbeda-beda, maka sebaiknya perusahaan tidak memperlakukan mereka dengan perlakuan yang sama. Pelanggan tidak hanya memiliki kebutuhan, preferensi, dan harapan yang berbeda, namun juga profil pendapatan dan biaya yang berbeda, dan karenanya harus dikelola dengan cara yang berbeda pula. Perlakuan yang sama terhadap semua pelanggan akan menyebabkan pelanggan yang tidak begitu bernilai bagi perusahaan akan berakhir menjadi penghancur nilai daripada pencipta nilai bagi perusahaan (Pramudiansyah, 2021).

Segmentasi pelanggan diperlukan untuk mengelompokkan pelanggan berdasarkan kesamaan karakteristik, yang membantu dalam menentukan strategi pemasaran yang tepat. Penggunaan teknik *data mining* merupakan solusi untuk segmentasi pelanggan, memungkinkan perusahaan mengelola data menjadi pengetahuan baru yang berharga. *Data mining* mengekstrak pengetahuan berharga dari data yang disimpan dalam basis data, data *warehouse*, atau tempat penyimpanan lainnya, sehingga mendukung pengambilan keputusan yang tepat. (Adiana et al., 2018).

Penelitian sebelumnya oleh Setiawan, Husni Amani, dan Wawan Tripiawan (2021) membahas segmentasi pelanggan menggunakan *Clustering K-Means* dan model RFM pada klinik kecantikan *Seoul Secret*. Tujuannya adalah mengelompokkan pelanggan berdasarkan karakteristik dan memberikan strategi pemasaran yang sesuai. *Clustering K-Means* digunakan untuk mengelompokkan pelanggan berdasarkan kesamaan karakteristik, sementara model RFM digunakan untuk mengelompokkan data berdasarkan atribut *Recency*, *Frequency*, dan *Monetary*. Dengan menggabungkan kedua metode tersebut, penelitian ini berhasil menggambarkan segmentasi pelanggan berdasarkan perilaku dan menghasilkan strategi pemasaran yang efektif.

Selain *Clustering K-Means* dan model RFM, analisis segmentasi pelanggan ini dapat menggunakan metode *Fuzzy C-Means*. Pada *Fuzzy C-Means* memperhitungkan dan mencari tingkat keanggotaan data dari dalam kelompok (Monalisa S et al., 2019), yang memungkinkan adanya overlap antara kelompok. Dengan *Fuzzy C-Means*, data pelanggan yang jauh dari pusat kluster masih dapat memiliki tingkat keanggotaan yang signifikan dalam kelompok tersebut (Ilham et al., 2020). Keunggulan *Fuzzy C-Means* dalam menangani tumpang tindih ini memungkinkan pemahaman yang lebih komprehensif tentang karakteristik dan perilaku pelanggan. .

Berdasarkan permasalahan diatas, akan dilakukan analisis data untuk segmentasi pelanggan dan penentuan pelanggan potensial guna mengatasi permasalahan yang ada. Data yang akan diteliti mencakup dua periode selama satu tahun, dari September 2020 hingga September 2022. Segmentasi pelanggan diperlukan untuk mengetahui pelanggan mana saja yang memberikan keuntungan terhadap perusahaan dan pelanggan mana saja yang memiliki potensi untuk kedepannya (Setiawan et al., 2021). Pada penelitian ini akan menggunakan RFM model sebagai penentu atribut dan dikombinasikan dengan *Fuzzy C-Means* untuk Klasterisasi. Model RFM digunakan untuk proses klasterisasi data pelanggan berdasarkan jumlah transaksi, nilai nominal transaksi, waktu transaksi (Basri et al., 2020).

## **1.2 Tujuan**

Tujuan dari penelitian ini adalah untuk melakukan segmentasi pelanggan yang berpotensi bagi perusahaan Hallogen Bogor. Dengan segmentasi yang tepat, perusahaan dapat memberikan perlakuan yang sesuai dan memaksimalkan keuntungan dari setiap pelanggan.

## **1.3 Ruang Lingkup**

Ruang lingkup penelitian adalah batasan-batasan yang dibuat untuk memfokuskan penelitian pada suatu area atau topik tertentu. Ruang lingkup penelitian ini meliputi :

1. Analisis perilaku konsumen pada Toko tas Hallogen Bogor
2. Penggunaan metode *Fuzzy C-Means* dan Model RFM
3. Pengumpulan data transaksi pelanggan Hallogen Bogor selama dua tahun dimulai dari September 2020 hingga September 2022.
4. Identifikasi kelompok pelanggan berdasarkan perilaku pembelian mereka

## **1.4 Manfaat Penelitian**

Manfaat penelitian adalah hasil yang diperoleh dari suatu penelitian yang bermanfaat bagi pihak-pihak yang terlibat dalam penelitian tersebut. Manfaat dari penelitian ini meliputi:

1. Penelitian ini membantu perusahaan mengidentifikasi pelanggan berpotensi dan menentukan strategi pemasaran yang tepat.
2. Hasil penelitian diharapkan dapat meningkatkan keuntungan perusahaan serta kepuasan pelanggan secara keseluruhan.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Tinjauan Pustaka**

##### **2.1.1 Pelanggan**

Pelanggan merupakan sumber keuntungan dan kehidupan bagi perusahaan, seperti halnya pada bisnis di bidang ritel. Pelanggan merupakan aset bagi perusahaan. Saat ini, perusahaan berlomba-lomba dalam memenangkan hati para pelanggan. Karena masing-masing pelanggan memiliki kebutuhan, preferensi, harapan, dan perilaku yang berbeda-beda, maka sebaiknya perusahaan tidak memperlakukan mereka dengan perlakuan yang sama. Pelanggan tidak hanya memiliki kebutuhan, preferensi, dan harapan yang berbeda, namun juga profil pendapatan dan biaya yang berbeda, dan karenanya harus dikelola dengan cara yang berbeda pula. Perlakuan yang sama terhadap semua pelanggan akan menyebabkan pelanggan yang tidak begitu bernilai bagi perusahaan akan berakhir menjadi penghancur nilai daripada pencipta nilai bagi perusahaan (Pramudiansyah, 2021).

##### **2.1.2 Segmentasi Pelanggan**

Segmentasi adalah salah satu konsep terpenting dalam pemasaran. Organisasi atau perusahaan jasa memiliki kemampuan yang berbeda untuk melayani jasa pelanggan yang berbeda. Perusahaan yang memahami kebutuhan pelanggan mungkin akan memilih untuk menerapkan pendekatan segmentasi berbasis kebutuhan, yang berfokus pada pelanggan yang menghargai atribut spesifik (Rukmana, 2020). Perusahaan mengimplementasikan segmentasi pelanggan berdasarkan fakta bahwa setiap pelanggan berbeda dan usaha pemasaran akan terlayani dengan baik jika terdapat target yang spesifik. Segmentasi umumnya digunakan untuk mendapatkan wawasan tentang sikap pelanggan, keinginan, pandangan, preferensi, opini tentang perusahaan dan kompetisi (Prasetyo et al., 2020).

##### **2.1.3 Model Recency, Frequency, Monetary**

Analisa model *Recency, Frequency, dan Monetary (RFM)* adalah proses menganalisis perilaku pelanggan. Hal ini umumnya digunakan dalam pemasaran database dan pemasaran langsung. Analisis RFM merupakan suatu Model perhitungan yang terdiri dari tiga Atribut *domain* yang memperhatikan transaksi pelanggan berdasarkan transaksi terakhir/keterkinian (*Recency*), jumlah transaksi (*Frequency*), serta nominal dari transaksi (*Monetary*). Tujuan dari RFM adalah untuk meramalkan perilaku Pelanggan di masa depan agar dapat mengarahkan keputusan segmentasi yang lebih baik. Model RFM ini merupakan metode yang sudah lama dan populer untuk mengukur hubungan dengan pelanggan (Pramudiansyah, 2021). Terdapat tiga parameter dalam model RFM yaitu *recency, frequency dan monetary*. Salah satu keuntungan dari model RFM antara lain mudah dipahami, efektif dalam mengidentifikasi pelanggan yang berharga. Lalu kelemahannya hanya dapat menggunakan sejumlah variabel seleksi yang terbatas (Wardani et al., 2022).

### 2.1.4 Clustering

*Clustering* merupakan proses partisi satu *set* objek data ke dalam himpunan bagian yang disebut dengan *cluster*. Objek yang di dalam *cluster* memiliki kemiripan karakteristik antar satu sama lainnya dan berbeda dengan *cluster* yang lain. Partisi tidak dilakukan secara manual melainkan dengan suatu algoritma *clustering*. Oleh karena itu, *clustering* sangat berguna dan bisa menemukan group atau kelompok yang tidak dikenal dalam data. *Clustering* banyak digunakan dalam berbagai aplikasi seperti misalnya pada *business intelligence*, pengenalan pola citra, *web search*, bidang ilmu biologi, dan untuk keamanan (*security*). Di dalam *business intelligence*, *clustering* bisa mengatur banyak *customer* ke dalam banyaknya kelompok (Annisa et al., n.d.).

### 2.1.5 Algoritma Fuzzy C-Means

Pengelompokan (*Clustering*) merupakan suatu teknik yang sudah cukup terkenal dan saat ini banyak diaplikasikan untuk pengelompokan data/objek ke dalam suatu kelompok data (*cluster*) sehingga setiap *cluster* mempunyai data yang mirip dan berbeda dengan data-data yang lainnya. Jika diberikan himpunan data yang berjumlah terhingga, yaitu  $X$ , maka permasalahan *clustering* dalam  $X$  adalah mencari beberapa pusat *cluster* yang dapat memberikan ciri kepada masing masing *cluster* dalam  $X$ . Contoh yang sangat sederhana dari proses pengelompokan data, dimana 3 kelompok di dalam suatu data dapat diidentifikasi dengan mudah. Pada kasus ini kriteria yang digunakan adalah jarak (dalam kasus ini jarak geometris). Adapun cara lain untuk melakukan pengelompokan adalah *conceptual clustering*. Dalam *conceptual clustering*, menurut konsep deskriptif objek didapat dari kecocokan tiap datanya (Yarif & WindartoS, 2018). Metode *Clustering* yang dipakai untuk segmentasi pelanggan pada penelitian ini adalah *Fuzzy C-Means (FCM)*. Metode ini merupakan metode non-hirarki dimana jumlah kelompoknya ditentukan terlebih dahulu berdasarkan kriteria yang diberikan. FCM menggunakan model pengelompokan *fuzzy* sehingga data dapat menjadi anggota dari semua kelas atau *cluster* terbentuk dengan derajat atau tingkat keanggotaan yang berbeda antara 0 hingga 1 (Ilham et al., 2020).

1. *Input data* yang akan di-*cluster*,  $X$ , berupa matriks berukuran  $n \times m$  ( $n$  = jumlah sampel data,  $m$  = atribut setiap data).  $X_{ij}$  data sampel ke- $i$  ( $i = 1, 2, \dots, n$ ), atribut ke- $j$  ( $j = 1, 2, \dots, m$ ).
2. Tentukan :
  - Jumlah *cluster* =  $c$
  - Pangkat =  $w$
  - *Maksimum* iterasi =  $MaxIter$
  - *Error* terkecil yang diharapkan =  $\xi$
  - Fungsi objektif awal =  $P_0 = 0$
  - Iterasi awal =  $t = 1$
3. Bangkitkan bilangan *random*  $\mu_{ik}$ ,  $i = 1, 2, \dots, n$ ;  $k = 1, 2, \dots, c$ ; sebagai elemen-elemen matriks partisi awal  $U$ . Lalu Hitung jumlah setiap kolom :

$$Q_i = \sum_{k=1}^c \mu_{ik} \text{ dengan } j = 1, 2, \dots, n \quad (1)$$

Hitung :

$$\mu_{ik} = \frac{\mu_{ik}}{Q_i} \quad (2)$$

4. Hitung pusat *Cluster* ke- $k$ ,  $V_{kj}$  dengan  $k = 1, 2, \dots, c$  dan  $j = 1, 2, \dots, m$

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \quad (3)$$

5. Hitung fungsi objektif pada iterasi ke- $t$ ,  $P_t$  :

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left( \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right) \quad (4)$$

6. Hitung perubahan matriks partisi :

$$\frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{\frac{-1}{w-1}}}{\sum_{k=1}^c [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{\frac{-1}{w-1}}} \quad (5)$$

dengan:  $i = 1, 2, \dots, n$  dan  $k = 1, 2, \dots, c$

7. Cek kondisi berhenti

- Jika:  $(|P_t - P_{t-1}| < \xi)$  atau  $(t > \text{MaxIter})$  maka berhenti;
- Jika tidak:  $t = t + 1$ , ulangi langkah ke-4.

### 2.1.6 *Elbow Method*

*Metode Elbow* digunakan untuk memilih jumlah *cluster* berdasarkan siku yang terbentuk pada suatu titik di grafik SSE dan didasarkan pada penurunan SSE yang besar. Jika nilai *cluster* sebelumnya ( $k-1$ ) dengan nilai *cluster* selanjutnya ( $k$ ) mengalami penurunan terbesar maka jumlah *cluster* tersebut yang tepat ( $k$ ). Metode ini menggunakan nilai *Sum of Square Error (SSE)* dari masing-masing jumlah *cluster*. Semakin besar jumlah *cluster*, maka SSE akan terus mengecil, sehingga jumlah *cluster* terbaik adalah jumlah *cluster* yang mengalami penurunan terbesar (Thamrin & Wijayanto, 2021). Rumus SSE dapat dituliskan sebagai :

$$SSE = \sum_{k=1}^K \sum X_i |X_i - C_k|^2 \quad (6)$$

Keterangan:

K : cluster.  
xi : data ke-i.  
ck : pusat cluster ke-k

### 2.1.7 Davies-Bouldin Index (DBI)

Prinsip pendekatan pengukuran DBI adalah memaksimalkan jarak antar kelompok (*inter-cluster*) dan meminimalkan jarak dalam kelompok (*intra-cluster*). Semakin kecil nilai DBI, semakin optimal skema pengelompokan (Gie & Jollyta, 2020). Berikut adalah empat tahapan dalam menghitung DBI (*Davis Bouldin Index*) untuk evaluasi hasil pengelompokan (Septiani et al., 2022) :

1. Menghitung *Sum Of Square Within Cluster (SSW)*, yang mengukur keterikatan anggota dalam satu kelompok atau seberapa mirip antara anggota dalam satu kelompok. Semakin kecil nilai SSW, semakin baik, karena itu menandakan tingkat kohesi yang tinggi dalam kelompok.

$$SSWi = \frac{1}{mi} * \sum_{j=i}^{mi} d(xj, ci)^2 \quad (7)$$

Keterangan:

mi : jumlah data dalam kelompok ke-i  
x : data dalam kelompok  
d(xj, ci) : jarak data ke centroid  
xj : data dalam kelompok tersebut  
ci : centroid kelompok ke-i

2. Menghitung *Sum Of Square Between Cluster (SSB)*, yang mengukur jarak antara kelompok yang cukup besar sehingga terpisah ke dalam kelompok lain. SSB bertujuan untuk mengetahui tingkat separasi atau heterogenitas antara kelompok.

$$SSBij = d(ci, cj) \quad (8)$$

Keterangan:

Ci = Kelompok 1  
Cj = Kelompok lainnya  
d(ci, cj) = Jarak antara centroid satu kelompok dengan kelompok lainnya.

3. Menghitung Rasio (Ri,j), yang berfungsi untuk mengetahui seberapa baik nilai perbandingan antara satu kelompok dengan kelompok lainnya. Jumlah kohesi (SSWi) harus kecil, sedangkan jumlah separasi (SSBij) harus lebih besar.

$$(Ri,j): Ri,j = \frac{(SSWi + SSWj)}{SSBij} \quad (9)$$

4. Menghitung DBI (*Davis Bouldin Index*), yang merupakan hasil akhir evaluasi yang menggabungkan nilai rasio (Ri,j) dari semua pasangan kelompok.

$$DBI = \frac{1}{n} * \sum \max * (Ri,j) \quad (10)$$

## 2.2 Penelitian Terdahulu

**Perancangan Segmentasi Pelanggan dengan Metode *Clustering KMeans* dan Model RFM pada Klinik Kecantikan *Seoul Secret* oleh Setiawan, Husni Amani, Wawan Tripiawan (2021).** Tujuan penelitian ini yaitu melakukan segmentasi pelanggan pengguna paket dan memberikan usulan strategi pemasaran. Penelitian ini menggunakan metode *Clustering K-Means* dan penerapan model RFM. *Clustering K-Means* digunakan merupakan salah satu metode *data mining* untuk mengelompokkan pelanggan berdasarkan karakteristik kesamaan pelanggan kedalam sejumlah K atau *cluster* yang telah ditentukan. Sedangkan RFM merupakan metode untuk mengelompokkan data berdasarkan atribut *Recency*, *Frequency* dan *Monetary*. Dengan mengkombinasi model RFM dan *Clustering* Algoritma *K-Means* dapat menggambarkan segmentasi pelanggan berdasarkan perilaku pelanggan. Kemudian penelitian ini juga menghasilkan strategi pemasaran yang dapat diberikan oleh pihak *Seoul Secret* diantaranya memberikan *cashback*, memberikan *voucher*, adanya *program membership*, penerapan sistem poin, penawaran paket *treatment* ditambah layanan pijat gratis, memberikan bonus *treatment*, menggunakan sistem *appointment*, penambahan dokter kecantikan dan adanya layanan prioritas.

**Implementasi Algoritma *Fuzzy C-Means (FCM)* dalam Pengklasterisasian Nilai Hidup Pelanggan dengan Model LRFM pada *Barbershop Omar Jalan Delima Pekanbaru Riau* oleh Waroka et al, (2020).** Penelitian ini dimulai dengan analisis data mentah (*cleaning*) yang kemudian ditransformasikan ke dalam model LRFM (*Length, Recency, Frequency, dan Monetary*) dan dinormalisasi dengan *range* nilai 0-1. Setelah itu, data di-klasterisasi menggunakan algoritma Fuzzy C-Means (FCM), dan nilai CLV per klaster dihitung. Hasilnya menunjukkan bahwa klaster dengan nilai CLV tertinggi adalah klaster ke-1 dengan nilai 0,355, diikuti oleh klaster ke-3 dengan nilai 0,314. Sedangkan klaster dengan nilai CLV terendah adalah klaster ke-2 dengan nilai 0,041, yang berada di bawah nilai rata-rata CLV sebesar 0,236. Klaster dengan nilai CLV tertinggi dapat dianggap sebagai kelompok pelanggan dengan loyalitas dan profitabilitas yang tinggi yang perlu dipertahankan.

**Sistem Segmentasi Loyalitas Pelanggan Berbasis RFM Menggunakan *K-Medoids Clustering* pada *Jayagiri Outdoor* oleh Ferry Rizki Setiawan, Yulison H. Chrissanto, Puspita N. Sabrina(2021).** Penelitian ini bertujuan untuk mengidentifikasi karakteristik pelanggan *Jayagiri Outdoor* dengan menggunakan konsep RFM (*Recency Frequency Monetary*) dan algoritma *K-Medoids Clustering*. Data transaksi pembelian pelanggan diolah menggunakan konsep RFM, yaitu data transaksi terakhir dimasukkan ke dalam *Recency*, total transaksi dimasukkan ke dalam *Frequency*, dan jumlah biaya yang dikeluarkan pelanggan dimasukkan ke dalam *Monetary*. Setelah itu, algoritma *K-Medoids Clustering* digunakan untuk mengelompokkan pelanggan berdasarkan karakteristik yang sama. Hasilnya terdapat 3 kelompok pelanggan, yaitu loyal, sedang, dan tidak loyal, dengan nilai *Recency, Frequency, dan Monetary* yang berbeda-beda pada setiap kelompok. Segmen loyal dianggap sebagai kelompok pelanggan dengan loyalitas tinggi yang perlu diperhatikan dan dikembangkan strategi pemasarannya.

### 2.3 Tabel Perbandingan

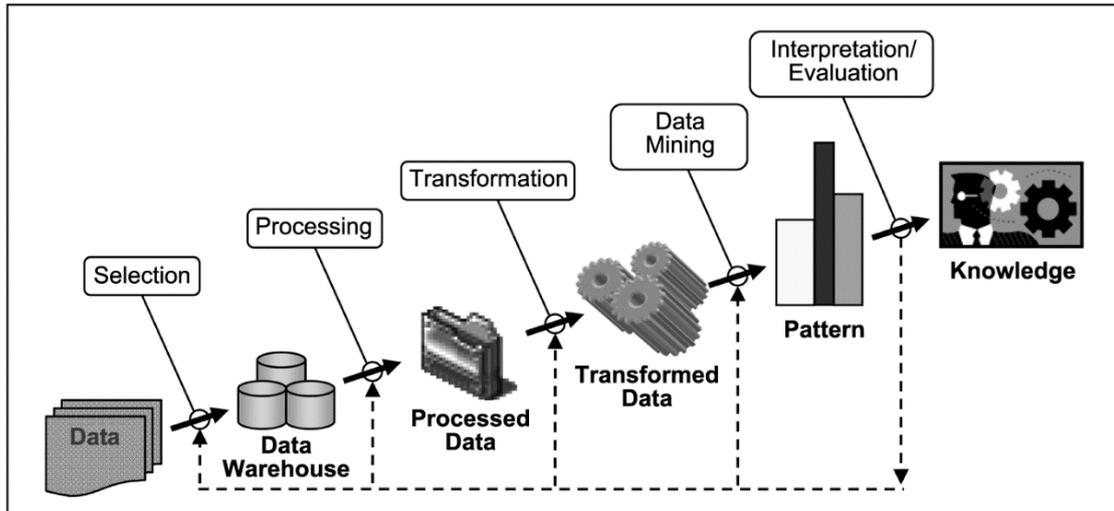
**Tabel 1.** Tabel Perbandingan

No	Nama	Objek Penelitian	Metode			Model	
			<i>K-Means</i>	<i>Fuzzy C-Means</i>	<i>K-Medoids</i>	RFM	LRFM
1	Setiawan, Husni Amani, Wawan Tripiawan. (2021)	KLINIK KECANTIKAN <i>SEOUL SECRET</i>	✓			✓	
2	Elison et al, (2020)	<i>BARBERSHOP</i> OMAR JALAN DELIMA PEKANBARU RIAU		✓			✓
3	Adha et al, (2022)	JAYAGIRI <i>OUTDOOR</i>			✓	✓	
4	Yusuf Fadilah Rukmana (2022)	TOKO TAS HALLOGEN BOGOR		✓		✓	

## BAB III METODOLOGI PENELITIAN

### 3.1 Metode Penelitian

Pada penelitian ini dilakukan pada sebuah Toko Hallogen bogor, penelitian ini menggunakan *Fuzzy C-Means* dan *Model RFM (Recency, Frequency, dan Monetary)* menggunakan metode *Knowledge Discovery in Databases (KDD)*. Tahapannya meliputi (Tahapan KDD) :



**Gambar 1.** *Knowledge Discovery in Database (KDD)*

#### 3.1.1 Selection

Pemilihan data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam *KDD (Knowledge Discovery in Database)*. Data hasil seleksi akan digunakan dalam proses data mining.

#### 3.1.2 Pre-processing

Proses *cleaning* mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data.

#### 3.1.3 Transformation

Proses *transformation* ini adalah proses di mana data yang telah dipilih akan diubah ke dalam bentuk di mana data dapat diproses dalam *data mining*.

#### 3.1.4 Mining

*Data mining* adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Pada tahap *data mining* segmentasi pelanggan, data pelanggan yang telah dipersiapkan akan dianalisis menggunakan metode *Fuzzy C-Means clustering*.

#### 3.1.5 Interpretation / Evaluation

Pola informasi yang dihasilkan dari proses *data mining* perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan.

### **3.1.6 Knowledge**

Tahap terakhir dari KDD adalah menyajikan pengetahuan yang baru ditemukan dengan cara yang mudah dimengerti dan berguna bagi pemangku kepentingan.

### **3.2 Waktu dan Tempat Penelitian**

Penelitian ini dilaksanakan pada bulan September yang bertempat di Laboratorium Ilmu Komputer Gedung FMIPA 2, Universitas Pakuan Jl. Pakuan, RT.02/RW.06 Tegallega, Kecamatan Bogor Tengah, Kota Bogor, Jawa Barat 16129.

### **3.3 Alat dan Bahan**

#### **3.3.1 Alat Penelitian**

Alat yang digunakan untuk penelitian ini dibagi menjadi dua yaitu Perangkat Lunak (*Software*) dan Perangkat Keras (*Hardware*).

##### **a. Perangkat Lunak (*Software*)**

1. *Operating Windows* 10 64 bit.
2. Google Collabs.
3. Visual Studio Code.
4. *Web Browser* Google Chrome dan Mozilla Firefox.
5. Microsoft Word 2010 dan Microsoft Excel 2010.
6. Draw.io.

##### **b. Perangkat Keras (*Hardware*)**

1. Laptop LENOVO V110.
2. RAM 4096MB / 4.0 GB.
3. *Processor* AMD A9-9420 RADDEON R5,5 (2CPUs) 3.0 Ghz.
4. *Keyboard* dan *Mouse*

#### **3.3.2 Bahan Penelitian**

Bahan-bahan yang digunakan sebagai sumber informasi pada penelitian ini diantaranya :

1. Buku Panduan Skripsi
2. Jurnal-jurnal Penelitian
3. Data Penjualan Toko Hallogen

## BAB IV PERANCANGAN DAN IMPLEMENTASI

### 4.1 Perancangan

Tahap perancangan dalam *Knowledge Discovery in Databases* (KDD) bertujuan untuk merancang desain atau rencana yang mencakup langkah-langkah pengolahan data dalam analisis. Pada tahap ini, strategi dan pendekatan yang sesuai dengan tujuan penelitian harus direncanakan. Sumber data yang relevan dipilih, dan langkah-langkah *preprocessing* data diatur. Teknik transformasi data, seperti pembuatan model RFM dan normalisasi skor RFM, ditentukan. Algoritma atau metode, seperti algoritma *Fuzzy C-Means*, dipilih untuk tahap *mining* dalam segmentasi pelanggan. Metrik evaluasi yang sesuai dipilih, dan interpretasi serta penyajian hasil segmentasi juga dipertimbangkan. Semua langkah dan keputusan tersebut didokumentasikan secara rinci dalam laporan skripsi untuk memberikan dasar yang kuat bagi penelitian.

### 4.2 Implementasi

Tahapan implementasi *KDD (Knowledge Discovery in Databases)* adalah tahap penting dalam proses analisis data karena pada tahap ini, hasil yang telah dirancang pada tahap perancangan akan diuji dan dinilai keefektifannya. Berikut adalah penjelasan untuk setiap tahapannya:

#### 4.2.1 Data Selection

*Data selection* adalah proses memilih data yang relevan untuk dianalisis dari suatu kumpulan data yang lebih besar. Pada kasus data pelanggan yang mencakup 2 periode dengan rentang tanggal September hingga September, tahun 2020 hingga 2022, *data selection* dapat dilakukan dengan memilih data pelanggan yang memenuhi kriteria tertentu seperti periode data, jenis pelanggan, variabel kunci, dan rentang tanggal. Kriteria-kriteria tersebut digunakan untuk memilih data pelanggan yang lebih spesifik dan relevan sehingga analisis yang dilakukan dapat lebih fokus dan akurat. Berikut adalah tabel data pelanggannya.

index	No	No. Pesanan	Id_Username	Tanggal	Total_Item	Sales	Harga_Produk	Total_Pembayaran	Total
0	1	200901M3FQKR09	dezsyf	2020-09-01 00:00:00	1	135000	135000	135000	135000
1	2	200901J4GQ0GS6	dezsyf	2020-09-02 00:00:00	3	135000	135000	405000	405000
2	3	200902B33JKAME	sausanbatarlie	2020-09-02 00:00:00	1	29000	29000	29000	29000
3	4	200907M3390NKQ	sausanbatarlie	2020-09-03 00:00:00	2	29000	29000	58000	58000
4	5	200902N26UFBJB	sausanbatarlie	2020-09-05 00:00:00	4	29000	29000	116000	116000
5	6	20090994TTWDNA	dessyintan00	2020-09-09 00:00:00	1	99000	99000	99000	99000
6	7	20090995BRCH9D	teger.ayu	2020-09-09 00:00:00	1	399000	399000	399000	399000
7	8	200909ABARQGXG	larasaliart_23	2020-09-09 00:00:00	1	199000	199000	199000	199000
8	9	200910BR80S9J2	pustakahaura	2020-09-10 00:00:00	1	99000	99000	99000	99000
9	10	200911E2G2AY84	teger.ayu	2020-09-11 00:00:00	1	135000	135000	135000	135000
10	11	200913MCC6CBUQ	sekarbeauty.id	2020-09-13 00:00:00	1	99000	99000	99000	99000
11	12	200914MQAHE3T6	virdaazmin	2020-09-13 00:00:00	1	325000	325000	325000	325000
12	13	2009194UMQXE8T	mininecha	2020-09-19 00:00:00	1	99000	99000	99000	99000
13	14	2009195BEDRFT4	nayla.w.p	2020-09-19 00:00:00	1	150000	150000	150000	150000
14	15	2009203E81DKQE	ismaaaaa0	2020-09-20 00:00:00	1	29000	29000	29000	29000
15	16	2009208R75AHNQ	ismaaaaa0	2020-09-20 00:00:00	1	29000	29000	29000	29000
16	17	200929VW2OR9D	devita0.0	2020-09-29 00:00:00	1	29000	29000	29000	29000
17	18	2009292N5B6JG3	devita0.0	2020-09-29 00:00:00	1	29000	29000	29000	29000
18	19	2009293HHQ29FG	devita0.0	2020-09-29 00:00:00	1	29000	29000	29000	29000
19	20	20100133HJLW23	pustakahaura	2020-10-01 00:00:00	1	29000	29000	29000	29000
20	21	20100166V7B4UF	pustakahaura	2020-10-01 00:00:00	1	29000	29000	29000	29000
21	22	2010029E0CN6AS	aminsrw	2020-10-02 00:00:00	1	29000	29000	29000	29000
22	23	201006K12JTA5	nenengnurmeida	2020-10-06 00:00:00	1	29000	29000	29000	29000
23	24	201006M31DD024	juairiahdjamil	2020-10-06 00:00:00	1	29000	29000	29000	29000
24	25	201006MOD29NJV	juairiahdjamil	2020-10-06 00:00:00	1	29000	29000	29000	29000
25	26	20101013YQ89F4	lola_sagita19	2020-10-10 00:00:00	1	349000	349000	349000	349000

**Gambar 2.** Data Pelanggan yang digunakan

#### 4.2.2 Data Preprocessing

Tahapan pertama dalam *data preprocessing* adalah *cleaning*, yaitu Tahap ini melibatkan pemilihan data yang akan digunakan untuk analisis segmentasi. Data yang umumnya digunakan untuk analisis RFM termasuk data transaksi, tanggal pembelian, frekuensi pembelian, dan nilai transaksi. Dapat dilihat pada tabel dibawah ini.

**Tabel 2.** *Data Cleaning*

<b>Id_Username</b>	<b>Tanggal</b>	<b>Total_Item</b>	<b>Harga_Produk</b>
dezsyf	01/09/2020	1	Rp135.000
dezsyf	02/09/2020	3	Rp135.000
sausanbatarfie	02/09/2020	1	Rp29.000
sausanbatarfie	03/09/2020	2	Rp29.000
sausanbatarfie	05/09/2020	4	Rp29.000
dessyintan00	09/09/2020	1	Rp99.000
teger.ayu	09/09/2020	1	Rp399.000
larasatiart_23	09/09/2020	1	Rp199.000
pustakahaura	10/09/2020	1	Rp99.000
teger.ayu	11/09/2020	1	Rp135.000
sekarbeauty.id	13/09/2020	1	Rp99.000
virdaazmin	13/09/2020	1	Rp325.000
mininecha	19/09/2020	1	Rp99.000
nayla.w.p	19/09/2020	1	Rp150.000
ismaaaaa0	20/09/2020	1	Rp29.000
ismaaaaa0	20/09/2020	1	Rp29.000
devita0.0	29/09/2020	1	Rp29.000
devita0.0	29/09/2020	1	Rp29.000
devita0.0	29/09/2020	1	Rp29.000
pustakahaura	01/10/2020	1	Rp29.000
pustakahaura	01/10/2020	1	Rp29.000
aminsrw	02/10/2020	1	Rp29.000
nenengnurmeida	06/10/2020	1	Rp29.000

#### 4.2.3 Data Transformation

Pada tahap transformasi, terdapat dua proses yang dilakukan, yaitu mengubah data yang telah dipilih dan diproses pada tahap sebelumnya menjadi *RFM score* dan melakukan normalisasi data. Berikut adalah penjelasan singkat mengenai kedua proses tersebut:

##### 4.2.3.1 Menentukan *Recency*

Langkah berikutnya ialah melakukan pengelompokan data dengan memperhatikan kolom tanggal pada transaksi terakhir, dengan mengelompokkannya berdasarkan Id\_Username.

**Tabel 3. Recency**

Id_Username	Today	LastPurchaseDate	Recency
09035koc6e	2022-09-28	2022-09-14	14
0albi_	2022-09-28	2022-09-06	22
17340_	2022-09-28	2021-07-25	430
1904faris	2022-09-28	2021-07-11	444
211209.ela	2022-09-28	2021-11-24	308
4412r13c7l	2022-09-28	2022-09-14	14
57jkiahdyov	2022-09-28	2021-06-10	475
____syarihdy	2022-09-28	2022-05-24	127
_bopung.id	2022-09-28	2021-05-15	501
_refaariani	2022-09-28	2022-08-09	50
_vivievanessa	2022-09-28	2021-07-07	448
aaillam17	2022-09-28	2022-07-13	77
aangandiana	2022-09-28	2022-05-30	121
abdi.putri	2022-09-28	2021-05-07	509
abia.homewear	2022-09-28	2022-08-09	50

$$\text{Recency (09035koc6e)} = \text{Today} - \text{Last Purchase Date} \quad (11)$$

$$\text{Recency (09035koc6e)} = 2022-09-28 - 2022-09-14 = 14 \text{ Hari}$$

09035koc6e	= (2022-09-28) - (2022-09-14) =	14	Hari
0albi_	= (2022-09-28) - (2022-09-06) =	22	Hari
17340_	= (2022-09-28) - (2021-07-25) =	430	Hari
1904faris	= (2022-09-28) - (2021-07-11) =	444	Hari
211209.ela	= (2022-09-28) - (2021-11-24) =	308	Hari
4412r13c7l	= (2022-09-28) - (2022-09-14) =	14	Hari
57jkiahdyov	= (2022-09-28) - (2021-06-10) =	475	Hari
____syarihdy	= (2022-09-28) - (2022-05-24) =	127	Hari
_bopung.id	= (2022-09-28) - (2021-05-15) =	501	Hari
_refaariani	= (2022-09-28) - (2022-08-09) =	50	Hari
_vivievanessa	= (2022-09-28) - (2021-07-07) =	448	Hari
aaillam17	= (2022-09-28) - (2022-07-13) =	77	Hari
aangandiana	= (2022-09-28) - (2022-05-30) =	121	Hari
abdi.putri	= (2022-09-28) - (2021-05-07) =	509	Hari
abia.homewear	= (2022-09-28) - (2022-08-09) =	50	Hari

#### 4.2.3.2 Menentukan *Frequency*

Pada tahap ini, dilakukan perhitungan jumlah transaksi yang telah dilakukan oleh pelanggan dalam periode waktu tertentu. Dalam hal ini, ingin diketahui seberapa sering pelanggan melakukan transaksi selama periode tersebut.

**Tabel 4. Frequency**

Id_Username	LastPurchaseDate	Total Transaksi (F)
09035koc6e	2022-09-14	12
0albi_	2022-09-06	10
17340_	2021-07-25	2
1904faris	2021-07-11	2
211209.ela	2021-11-24	4
4412r13c7l	2022-09-14	5
57jkiahdyov	2021-06-10	2
___syarihdy	2022-05-24	2
_bopung.id	2021-05-15	2
_refaariani	2022-08-09	3
_vivievanessa	2021-07-07	5
aaillam17	2022-07-13	2
aangandiana	2022-05-30	1
abdi.putri	2021-05-07	6
abia.homewear	2022-08-09	6

**4.2.3.3 Menentukan Monetary**

Pada tahap ini, akan dihitung jumlah uang yang telah dikeluarkan oleh masing-masing pelanggan dalam jangka waktu tertentu. Ini berguna untuk mengetahui berapa banyak uang yang telah dihabiskan oleh pelanggan dalam periode tersebut.

**Tabel 5. Monetary**

Id_Username	Uang Yang Dikeluarkan (M)
09035koc6e	372600
0albi_	1286600
17340_	648000
1904faris	898000
211209.ela	954000
4412r13c7l	492600
57jkiahdyov	548000
___syarihdy	62000
_bopung.id	798000
_refaariani	99000
_vivievanessa	1511000
aaillam17	231000
aangandiana	31000
abdi.putri	1894000
abia.homewear	424000

#### 4.2.3.4 Menentukan RFM Score

Skor RFM digunakan untuk mengklasifikasikan pelanggan berdasarkan tiga faktor utama: Recency, Frequency, dan Monetary. Rentang skor RFM adalah 1 hingga 4, dengan skor 4 sebagai skor tertinggi yang menunjukkan nilai pelanggan yang paling berharga atau berpotensi tinggi bagi perusahaan. Semakin tinggi skor RFM, semakin aktif, sering berinteraksi, dan mengeluarkan uang lebih banyak pelanggan dalam hubungannya dengan perusahaan. Sebaliknya, skor RFM 1 menunjukkan ketidakaktifan, jarang berinteraksi, dan pengeluaran yang minim. Skor RFM 2 menunjukkan tingkat aktivitas, frekuensi, dan nilai yang sedang, sementara skor RFM 3 menunjukkan tingkat aktivitas, frekuensi, dan nilai yang cukup tinggi. Dengan memperhatikan skor RFM, perusahaan dapat mengidentifikasi segmen pelanggan yang perlu mendapatkan perhatian lebih dan mengarahkan upaya pemasaran untuk mempertahankan dan meningkatkan hubungan dengan mereka.

**Tabel 6.** RFM Score

Id_Username	Transaksi Akhir (Hari)	Total Transaksi	Uang Yang Dikeluarkan	R_Score	F_Score	M_Score
09035koc6e	14	12	372600	4	4	3
0albi_	22	10	1286600	4	4	4
17340_	430	2	648000	1	2	4
1904faris	444	2	898000	1	2	4
211209.ela	308	4	954000	1	3	4
4412r13c7l	14	5	492600	4	3	3
57jkiahiov	475	2	548000	1	2	4
____syarihdy	127	2	62000	2	2	1
_bopung.id	501	2	798000	1	2	4
_refaariani	50	3	99000	3	3	1
_vivievanessa	448	5	1511000	1	3	4
aaillam17	77	2	231000	3	2	2
aangandiana	121	1	31000	2	1	1
abdi.putri	509	6	1894000	1	4	4
abia.homewear	50	6	424000	3	4	3

#### 4.2.3.5 Mengubah Menjadi Data Normalisasi

Tahap selanjutnya adalah tahap normalisasi data, Salah satu metode yang sering digunakan dalam normalisasi data adalah *Min-Max Scaler*. Metode ini bertujuan untuk mengubah skala variabel menjadi nilai antara 0 dan 1. Proses normalisasi dilakukan dengan menggunakan metode *Min-Max Scaler* yang mana dilakukan dengan mengurangi nilai setiap variabel dengan nilai *minimum* variabel tersebut, kemudian hasilnya dibagi dengan selisih nilai *maksimum* dan *minimum* variabel tersebut.

$$X_{new} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (12)$$

Keterangan :

$X_{new}$  adalah nilai baru setelah dilakukan normalisasi

$X$  adalah nilai awal dari variabel

$X_{min}$  adalah nilai minimum dari variabel

$X_{max}$  adalah nilai maksimum dari variabel

$$09035koc6e (R_{Norm}) = \frac{4 - 1}{4 - 1} = 1$$

$$09035koc6e (F_{Norm}) = \frac{4 - 1}{4 - 1} = 1$$

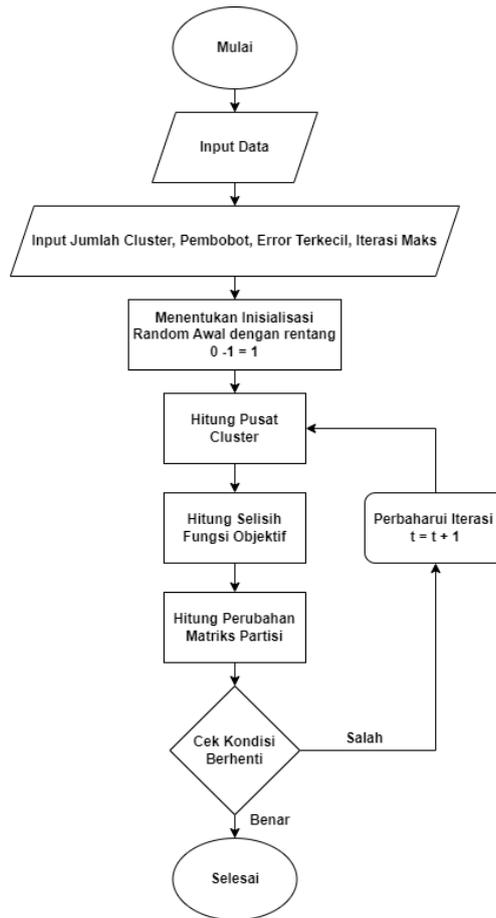
$$09035koc6e (M_{Norm}) = \frac{3 - 1}{4 - 1} = 0,67$$

**Tabel 7.** Normalisasi *RFM Score*

<b>Id_Username</b>	<b>R_Score</b>	<b>F_Score</b>	<b>M_Score</b>	<b>R</b>	<b>F</b>	<b>M</b>
09035koc6e	4	4	3	1.0	1.0	0.67
0albi_	4	4	4	1.0	1.0	1.0
17340_	1	2	4	0.0	0.33	1.0
1904faris	1	2	4	0.0	0.33	1.0
211209.ela	1	3	4	0.0	0.67	1.0
4412r13c7l	4	3	3	1.0	0.67	0.67
57jkiahdyov	1	2	4	0.0	0.33	1.0
____syarihdy	2	2	1	0.33	0.33	0.0
_bopung.id	1	2	4	0.0	0.33	1.0
_refaariani	3	3	1	0.67	0.67	0.0
_vivievanessa	1	3	4	0.0	0.67	1.0
aailam17	3	2	2	0.67	0.33	0.33
aangandiana	2	1	1	0.33	0.0	0.0
abdi.putri	1	4	4	0.0	1.0	1.0
abia.homewear	3	4	3	0.67	1.0	0.67

#### 4.2.4 Metode *Fuzzy C-Means*

Setelah tahap *preprocessing*, data pelanggan akan dianalisis menggunakan metode *Fuzzy C-Means* clustering dalam bahasa pemrograman Python. Metode ini memungkinkan pengelompokan pelanggan berdasarkan kesamaan atribut seperti *Recency*, *Frequency*, dan *Monetary*. Dengan menggunakan metode ini, kita dapat memahami perilaku pelanggan dengan lebih mendalam dan mengidentifikasi kelompok pelanggan potensial, sehingga memungkinkan perusahaan untuk merancang strategi pemasaran yang lebih efektif dan meningkatkan pendapatan. Berikut adalah *flowchart* proses *clustering* menggunakan *Fuzzy C-Means*.



**Gambar 3.** Algoritma *Fuzzy C-Means*

Berikut merupakan langkah-langkah implementasi menggunakan Algoritma *Fuzzy C-Means* :

#### 4.2.4.1 Penentuan Jumlah Cluster

Tahap ini merupakan tahap awal dari proses *clusterisasi* pada metode *Fuzzy C-Means*. Pada tahap ini, pengguna dapat menentukan jumlah *cluster* secara bebas sesuai keinginan, serta menentukan parameter lainnya seperti pangkat (*fuzzy exponent*), jumlah iterasi *maximum* (*maxiter*), dan toleransi *error* terkecil (*tol*). Penentuan parameter ini dapat memengaruhi hasil *clustering* dan harus dipilih dengan hati-hati untuk mengoptimalkan kinerja *clustering*.

**Tabel 8.** Tabel Penentuan *Cluster*

Jumlah <i>Cluster</i> (C)	3
Pangkat	2
Max Iterasi	100
Error Terkecil	0,01

#### 4.2.4.2 Inisialisasi *Random Awal*

Proses penentuan inisialisasi nilai *random* awal pada *Fuzzy C-Means (FCM)* dilakukan dengan menghitung jarak *euclidean* antara setiap objek data dengan nilai *centroid* awal yang dihasilkan secara acak. Nilai derajat keanggotaan pada *Fuzzy C-Means (FCM)* harus berada di dalam rentang 0 hingga 1, karena derajat keanggotaan merupakan representasi dari tingkat keterkaitan suatu objek data dengan setiap *cluster* yang ada. Nilai derajat keanggotaan 0 menunjukkan bahwa suatu objek data tidak memiliki keterkaitan dengan suatu *cluster* tertentu, sedangkan nilai derajat keanggotaan 1 menunjukkan bahwa suatu objek data memiliki keterkaitan penuh dengan suatu *cluster* tertentu. Nilai derajat keanggotaan di antara 0 dan 1 menunjukkan tingkat keterkaitan suatu objek data dengan setiap *cluster* yang ada. Berikut tabel inisialisasi *random* awal pada data sampel.

**Tabel 9.** Inisialisasi *Random Awal*

Inisialisasi <i>Random Awal/ C</i>		
C0	C1	C2
0,21	0,29	0,5
0,11	0,14	0,74
0,02	0,15	0,82
0,02	0,41	0,56
0,23	0,15	0,62
0,22	0,26	0,52
0,01	0,31	0,68
0,03	0,12	0,85
0,19	0,39	0,42
0,19	0,34	0,46
0,02	0,41	0,56
0,22	0,26	0,52
0,11	0,14	0,74
0,03	0,12	0,85
0,19	0,39	0,42

#### 4.2.4.3 Perhitungan Pusat *Cluster*

Pada tahap selanjutnya adalah menghitung pusat *cluster*, pusat *cluster* didapat melalui proses perhitungan antara data parameter dengan nilai *random* keanggotaan yang menghasilkan sebuah nilai yang nantinya menjadi titik pusat *cluster*. Dapat dilihat pada tabel berikut:

Miu	=	0,3294	1,1844	6,0418
C0	=	0,189692	0,214005	0,205434
C1	=	0,407171	0,69251	0,889254
C2	=	1,928261	3,350941	4,148988

$$\mu_{ik} = \frac{\mu_{ik}}{Q_i} \quad (13)$$

Dimana :

$$\begin{array}{lll} C0 = \frac{0,189692}{0,3294} & C0 = \frac{0,214005}{0,3294} & C0 = \frac{0,205434}{0,3294} \\ C1 = \frac{0,407171}{1,1844} & C1 = \frac{0,69251}{1,1844} & C1 = \frac{0,889254}{1,1844} \\ C2 = \frac{1,928261}{6,0418} & C2 = \frac{3,350941}{6,0418} & C2 = \frac{4,148988}{6,0418} \end{array}$$

Dari perhitungan tersebut, dapat diketahui bahwa hasil pusat klaster adalah sebagai berikut:

**Tabel 10.** Pusat *Cluster* Awal

Pusat <i>Cluster</i>			
C0	0,575871	0,649681	0,623661
C1	0,343778	0,584693	0,750805
C2	0,319153	0,554626	0,686714

#### 4.2.4.4 Fungsi Objektif

Pada tahapan ini merupakan sebuah perhitungan untuk mencari nilai optimum, perhitungan kembali dilakukan setiap data untuk masuk pada setiap *cluster*, jika nilai yang dihasilkan belum optimal maka dilakukan iterasi selanjutnya untuk menentukan nilai yang optimal. Nilai fungsi objektif dapat dilihat pada tabel berikut:

**Tabel 11.** Fungsi Objektif Awal

Selisih Fungsi Objektif			
F0	F1	F2	FT
0,01344	0,05127	0,165547	0,230257
0,005375	0,013038	0,416208	0,434621
0,00023	0,005516	0,168412	0,174158
0,00023	0,04121	0,078546	0,119985
0,025057	0,00422	0,082	0,111277
0,00883	0,030044	0,129019	0,167893
5,75E-05	0,023559	0,115815	0,139431
0,000496	0,009054	0,377254	0,386804
0,020774	0,037287	0,044182	0,102243
0,014376	0,078308	0,128649	0,221333
0,000189	0,031529	0,066897	0,098615
0,009549	0,02355	0,081335	0,114434
0,010545	0,017753	0,426747	0,455045
0,000536	0,00508	0,287819	0,293435
0,004828	0,043414	0,056753	0,104995
<b>Total</b>			<b>3,154526</b>

#### 4.2.4.5 Cek Kondisi Berhenti

Pada tahapan ini menunjukkan bahwa nilai iterasi ke-1 masih belum optimal dikarenakan nilai fungsi objektif pada iterasi ke-1 masih belum optimum, maka dilakukan perulangan hingga mencapai nilai yang optimum.

##### a. Iterasi Awal

Hasil dari iterasi awal ini akan digunakan sebagai titik awal untuk proses iteratif selanjutnya. Selanjutnya, nilai keanggotaan untuk setiap data dan pusat kelompok (*centroid*) akan diperbarui pada setiap iterasi dengan menggunakan rumus-rumus matematis yang telah ditentukan. Proses ini akan berlanjut hingga nilai keanggotaan dan pusat kelompok konvergen atau sampai mencapai batas iterasi yang telah ditentukan sebelumnya. Dapat dilihat pada tabel berikut.

**Tabel 12.** Iterasi Awal

Selisih Fungsi Objektif	
P1 (Nilai Iterasi ke-1)	3,154526
P0 (Nilai Iterasi ke-0)	0
P1-P0 (Selisih Fungsi Objektif)	3,154526 - 0

Karena  $|P1 - P0| = |3,154526 - 0| = 3,154526 > \xi$  dalam hal ini Error terkecil = 0.01, dan iterasi = 1 < MaxIter(100), maka proses dilanjutkan ke iterasi selanjutnya (t=n).

##### b. Iterasi Akhir

Pada hasil iterasi akhir dalam *Fuzzy C-Means clustering*, nilai keanggotaan dan pusat kelompok telah konvergen dan mencapai nilai yang stabil. Proses iteratif pada FCM terus berlanjut hingga nilai keanggotaan dan pusat kelompok tidak berubah atau berubah sangat kecil dari iterasi sebelumnya.

**Tabel 13.** Iterasi Akhir

Selisih Fungsi Objektif	
P13 (Nilai Iterasi ke-13)	8,708984
P14 (Nilai Iterasi ke-14)	8,690167
P13-P14 (Selisih Fungsi Objektif)	8,708984 - 8,690167

Karena  $|P13 - P14| = |8,708984 - 8,690167| = 0,018817 < \xi$  Error terkecil = 0.01, dan iterasi = 14 < MaxIter(100), maka proses dihentikan.

##### c. Hasil Cluster

Dari data dibawah ini, dapat diketahui bahwa nilai *cluster* berada pada iterasi ke-14. Label *Cluster* 0 memiliki arti "*High value*", *cluster* 1 memiliki arti "*Mid value*", dan *cluster* 2 memiliki arti "*Low value*". Dapat dilihat pada tabel berikut.

**Tabel 14.** Tabel *Cluster*

Matriks Partisi U			
U0	U1	U2	Cluster
<b>0,902140978</b>	0,062075009	0,035784014	0
<b>0,737360795</b>	0,185883066	0,076756139	0
0,089619149	<b>0,744749151</b>	0,1656317	1
0,089619149	<b>0,744749151</b>	0,1656317	1
0,041600832	<b>0,909852495</b>	0,048546673	1
<b>0,93111072</b>	0,041308358	0,027580923	0
0,089619149	<b>0,744749151</b>	0,1656317	1
0,090298051	0,079703204	<b>0,829998745</b>	2
0,089619149	<b>0,744749151</b>	0,1656317	1
<b>0,439793498</b>	0,174340764	0,385865738	0
0,041600832	<b>0,909852495</b>	0,048546673	1
0,367380394	0,179401589	<b>0,453218017</b>	2
0,088345336	0,087231887	<b>0,824422777</b>	2
0,137145522	<b>0,753374391</b>	0,109480087	1
<b>0,779574486</b>	0,15512695	0,065298565	0

d. *Davies-Bouldin Index*

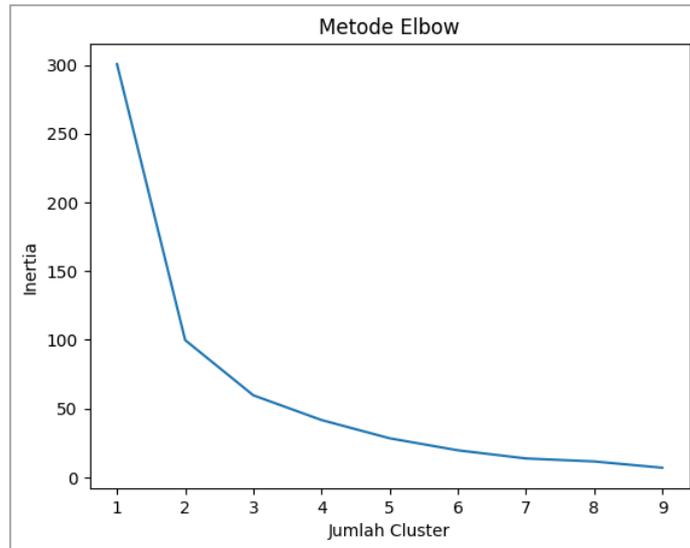
Hasil evaluasi menggunakan DBI dengan skor 0,83 pada data hasil pengelompokan dengan 3 *cluster* menunjukkan bahwa skema pengelompokan masih memiliki tingkat keberagaman atau heterogenitas antara kelompok yang cukup tinggi. Skor yang mendekati atau melebihi 1 pada DBI dianggap rendah, karena DBI memiliki rentang nilai dari 0 ke atas, dan semakin dekat ke 0, semakin baik. Oleh karena itu, skor 0,83 pada DBI menunjukkan bahwa masih ada ruang untuk meningkatkan kualitas pengelompokan untuk menghasilkan kelompok yang lebih homogen dan terpisah dengan jelas antara satu sama lain.

**4.2.5 Evaluasi**

Tahapan evaluasi pada FCM dapat dilakukan untuk memastikan hasil *clustering* yang optimal.

a. Evaluasi *Cluster* dengan *Elbow Method*

*Elbow Method* adalah salah satu metode evaluasi yang dapat digunakan untuk menentukan jumlah *cluster* yang optimal. Pada grafik *Elbow Method*, titik yang terletak di siku kurva menandakan jumlah *cluster* yang optimal. Dari grafik *elbow*, dapat diketahui bahwa jumlah *cluster* optimalnya adalah 2. Jika sudah tahu jumlah *cluster* optimalnya, lalu akan diteruskan dan mengulang proses pengelompokan (*clusterisasi*) data. Berikut hasil dari evaluasi *cluster* menggunakan *Elbow Method*.



**Gambar 4.** Evaluasi *Cluster Elbow Method*

Berdasarkan analisis, ditemukan bahwa jumlah kelompok pelanggan yang optimal adalah 2 kelompok. Hal ini didasarkan pada grafik *elbow* yang menunjukkan penurunan yang signifikan pada setiap kelompok tambahan dan menjadi lebih lambat setelahnya. Dengan demikian, kelompok pelanggan dapat dikelompokkan menjadi 2 kelompok berdasarkan perilaku pembelian mereka.

b. Hasil Perhitungan 2 *Cluster*

Dari hasil pengolahan data, dapat disimpulkan bahwa *clustering* telah diselesaikan. Terdapat 2 label *cluster* yang diberi makna bahwa *cluster 0* memiliki arti "Low Value" dan *cluster 1* memiliki arti "High Value".

**Tabel 15.** Hasil 2 *Cluster*

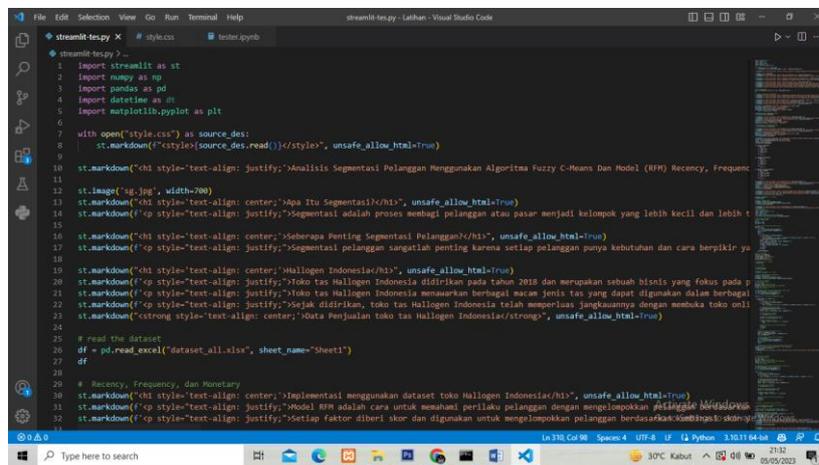
C0	C1	Cluster
0.38039734801297903	0.6196026519870209	High Value
0.32368878457851524	0.6763112154214848	High Value
0.1285977111470081	0.8714022888529919	High Value
0.1285977111470081	0.8714022888529919	High Value
0.009519053471599809	0.9904809465284001	High Value
0.4237619656826772	0.5762380343173228	High Value
0.1285977111470081	0.8714022888529919	High Value
0.8115061643800228	0.18849383561997718	Low Value
0.1285977111470081	0.8714022888529919	High Value
0.6129906162784754	0.38700938372152466	Low Value
0.009519053471599809	0.9904809465284001	High Value
0.6646668525659398	0.3353331474340601	Low Value
0.8828064098252886	0.11719359017471134	Low Value
0.10761195128133562	0.8923880487186644	High Value
0.31569789421824024	0.6843021057817598	High Value

#### 4.2.6 Knowledge

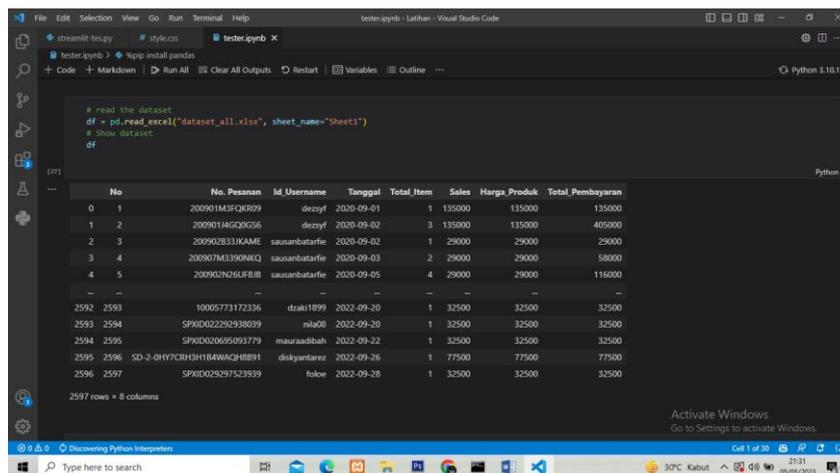
Tahap *Knowledge* dalam proses KDD merupakan tahap akhir dari seluruh proses. Tahap ini melibatkan interpretasi hasil analisis dan evaluasi kesesuaian dengan tujuan analisis. Hasil analisis kemudian dinilai berdasarkan kebenaran, konsistensi, dan relevansi.

### 4.3 Tahapan Implementasi Menggunakan Visual Studio Code

Implementasi *Visual Studio Code* dalam membuat kode program memiliki beberapa tahapan. Pertama, buka *Visual Studio Code* dan buatlah file baru dengan format sesuai kebutuhan. Kemudian, tuliskan kode program yang dibutuhkan dengan menggunakan bahasa pemrograman yang sesuai. Selanjutnya, lakukanlah *debugging* dan pengujian untuk memastikan program berjalan dengan baik dan sesuai dengan yang diinginkan. Terakhir, simpan kode program yang telah dibuat untuk digunakan pada masa yang akan datang atau untuk dikembangkan lebih lanjut.



Gambar 5. Implementasi *Visual Studio Code* pada Analisis Data



Gambar 6. Implementasi *Visual Studio Code* pada Proses Pengolahan Data

Kode diatas merupakan implementasi untuk melakukan *clustering* dengan menggunakan *Fuzzy C-Means*. Pada kode tersebut menunjukkan bahwa *clusterisasi* telah dilakukan atau sedang diproses.

## **BAB V**

### **HASIL DAN PEMBAHASAN**

#### **5.1 Hasil**

Berikut ini adalah hasil analisis menggunakan metode segmentasi *Fuzzy C-Means* yang diaplikasikan pada data pelanggan Toko Hallogen Bogor dengan bahasa pemrograman Python. Pada hasil ini, akan dijelaskan setiap proses yang dilakukan sebagaimana rancangan yang telah dirancang sebelumnya. Setelah proses *clustering* selesai dilakukan, hasil yang dihasilkan biasanya akan ditampilkan dalam bentuk visualisasi grafik atau tabel. Dalam hal ini, grafik yang dapat digunakan adalah grafik *scatter plot* yang menunjukkan lokasi dari setiap data dalam ruang atribut yang dianggap. Setiap *cluster* akan ditunjukkan dengan warna yang berbeda untuk mempermudah pemahaman.

##### **5.1.1 Tahap Pengumpulan Data**

Setelah melakukan studi literatur berdasarkan pokok bahasan segmentasi, dilakukanlah proses observasi dan pengamatan secara non-partisipan terhadap data pelanggan Toko Hallogen Bogor dari 2020 hingga 2022. Lalu melakukan pengajuan permohonan data pelanggan kepada pihak Toko Hallogen Bogor, yang akan digunakan dalam penelitian ini. Data pelanggan yang akan diambil meliputi data dari periode satu tahun September hingga September berikutnya, yaitu dari September 2020 hingga September 2022. Data tersebut akan digunakan dalam analisis segmentasi pelanggan menggunakan metode *Fuzzy C-Means*, dengan tujuan untuk memperoleh pemahaman yang lebih mendalam mengenai perilaku dan kebutuhan pelanggan Toko Hallogen Bogor, sehingga dapat diambil keputusan strategis yang lebih efektif dalam meningkatkan kepuasan pelanggan dan mengembangkan bisnis toko tersebut.

##### **5.1.2 Hasil Data Selection**

Dalam pemilihan data ini, berfokus pada data pelanggan toko Hallogen Bogor, karena data ini akan memberikan informasi yang relevan untuk memahami perilaku dan kebutuhan pelanggan toko tersebut. Dengan menggunakan data dari periode satu tahun September hingga September berikutnya, dan dapat memperoleh informasi yang lebih lengkap mengenai pola pembelian dan preferensi pelanggan selama satu siklus tahunan. Hal ini akan membantu untuk melakukan analisis segmentasi yang lebih akurat dan relevan untuk pengambilan keputusan bisnis yang lebih baik.

##### **5.1.3 Hasil Data Preprocessing**

Sebelum melakukan analisis segmentasi pelanggan menggunakan model *RFM* (*Recency, Frequency, Monetary*) dan Algoritma *Fuzzy C-Means*, akan dilakukan tahap *data preprocessing* terlebih dahulu. Sehingga data yang digunakan dalam analisis segmentasi pelanggan menjadi lebih akurat dan efektif.

#### 5.1.4 Hasil Data Transformasi

Pada tahap data transformasi dapat dilakukan dengan mengubah data transaksi pelanggan menjadi skor RFM. *RFM score* adalah skor yang diberikan untuk masing-masing pelanggan berdasarkan tiga faktor utama, yaitu *Recency*, *Frequency*, dan *Monetary*. Setelah mendapatkan skor RFM untuk setiap pelanggan, tahap selanjutnya adalah normalisasi skor RFM. Setelah skor RFM dinormalisasi, data pelanggan siap untuk dimasukkan ke dalam algoritma *Fuzzy C-Means*. Pada algoritma *Fuzzy C-Means*, data pelanggan akan dikelompokkan ke dalam beberapa *cluster* berdasarkan kemiripan atribut. *Cluster-cluster* ini akan membantu Toko Hallogen Bogor dalam memahami karakteristik dan preferensi pelanggan mereka dengan lebih baik.

#### 5.1.5 Hasil Analisis Segmentasi Pelanggan

Pada tahap ini, dilakukan perhitungan untuk melakukan segmentasi pelanggan dengan menggunakan dua model yaitu model RFM dan *Fuzzy C-Means*. Model RFM membagi pelanggan berdasarkan ketepatan waktu, frekuensi, dan nilai pembelian, sedangkan *Fuzzy C-Means* membagi pelanggan ke dalam kelompok-kelompok berdasarkan preferensi dan perilaku pembelian. Dengan segmentasi pelanggan, perusahaan dapat memahami preferensi dan kebutuhan pelanggan secara lebih detail dan mengembangkan strategi pemasaran yang lebih tepat sasaran. Dapat dilihat dibawah ini.

##### a. Model RFM

Fungsi *recency score* mengambil data *recency* dan menentukan skor berdasarkan empat kondisi. Skor 4 diberikan jika nilai *recency* kurang dari 50 hari, skor 3 jika nilai *recency* berada di antara 50 dan kurang dari 100 hari, skor 2 jika nilai *recency* berada di antara 100 dan kurang dari 200 hari, dan skor 1 jika nilai *recency* lebih dari 200 hari.

Fungsi *frequency score* mengambil data *frequency* dan menentukan skor berdasarkan empat kondisi. Skor 1 diberikan jika nilai *frequency* kurang dari atau sama dengan 1 kali transaksi, skor 2 diberikan jika nilai *frequency* berada di antara lebih dari 1 kali transaksi dan kurang dari atau sama dengan 2 kali transaksi, skor 3 diberikan jika nilai *frequency* berada di antara lebih dari 2 kali transaksi dan kurang dari 5 kali transaksi, dan skor 4 diberikan jika nilai *frequency* lebih dari 5 kali transaksi.

Fungsi *monetary score* mengambil data *monetary* dan menentukan skor berdasarkan empat kondisi. Skor 1 diberikan jika nilai *monetary* kurang dari atau sama dengan 205,000, skor 2 diberikan jika nilai *monetary* berada di antara lebih dari 205,000 dan kurang dari 351,000, skor 3 diberikan jika nilai *monetary* berada di Antara lebih dari 351,000 dan kurang dari 497,000, dan skor 4 diberikan jika nilai *monetary* lebih dari 497,000. Pengambilan data pelanggan sesuai model RFM dapat dilihat di tabel berikut :

**Tabel 16.** Tabel Model RFM

Id_Username	Transaksi Akhir (Hari)	Total Transaksi	Uang Yang Dikeluarkan	R_Score	F_Score	M_Score
09035koc6e	14	12	372600	4	4	3
0albi_	22	10	1286600	4	4	4
17340_	430	2	648000	1	2	4
1904faris	444	2	898000	1	2	4
211209.ela	308	4	954000	1	3	4
4412r13c7l	14	5	492600	4	3	3
57jkiahiov	475	2	548000	1	2	4
____syarihdy	127	2	62000	2	2	1
_bopung.id	501	2	798000	1	2	4
_refaariani	50	3	99000	3	3	1
_vivievanessa	448	5	1511000	1	3	4
aaillam17	77	2	231000	3	2	2
Aangandiana	121	1	31000	2	1	1
abdi.putri	509	6	1894000	1	4	4
abia.homewear	50	6	424000	3	4	3

	Id_Username	Recency	Frequency	Monetary	R	F	M
0	09035koc6e	14	12	372600	4	4	3
1	0albi_	22	10	1286600	4	4	4
2	17340_	430	2	648000	1	2	4
3	1904faris	444	2	898000	1	2	4
4	211209.ela	308	4	954000	1	3	4
...	...	...	...	...	...	...	...
930	zulfamarsalia	441	1	289000	1	1	2
931	zulfiaswinn	141	1	31000	2	1	1
932	zuroidadiana	449	1	349000	1	1	2
933	zyadakebuli	461	1	349000	1	1	2
934	zyahamida	444	3	873000	1	3	4

935 rows × 7 columns

**Gambar 7.** RFM Model Score

b. Perhitungan *Fuzzy C-Means*

Label hasil dari *Cluster Fuzzy C-Means* dapat dimasukkan pada data RFM yang sudah dinormalisasi seperti dibawah ini, Label kluster kemudian dapat ditambahkan sebagai kolom tambahan dalam kumpulan data RFM untuk analisis lebih lanjut.

	R	F	M	Cluster	C0	C1
0	1.000000	1.000000	0.666667	1	0.380397	0.619603
1	1.000000	1.000000	1.000000	1	0.323689	0.676311
2	0.000000	0.333333	1.000000	1	0.128598	0.871402
3	0.000000	0.333333	1.000000	1	0.128598	0.871402
4	0.000000	0.666667	1.000000	1	0.009519	0.990481
...	...	...	...	...	...	...
930	0.000000	0.000000	0.333333	0	0.988965	0.011035
931	0.333333	0.000000	0.000000	0	0.882806	0.117194
932	0.000000	0.000000	0.333333	0	0.988965	0.011035
933	0.000000	0.000000	0.333333	0	0.988965	0.011035
934	0.000000	0.666667	1.000000	1	0.009519	0.990481

935 rows × 6 columns

**Gambar 8.** Hasil Perhitungan dengan Algoritma *Fuzzy C-Means*

	R	F	M	Cluster
0	1.000000	1.000000	0.666667	High Value
1	1.000000	1.000000	1.000000	High Value
2	0.000000	0.333333	1.000000	High Value
3	0.000000	0.333333	1.000000	High Value
4	0.000000	0.666667	1.000000	High Value
...	...	...	...	...
930	0.000000	0.000000	0.333333	Low Value
931	0.333333	0.000000	0.000000	Low Value
932	0.000000	0.000000	0.333333	Low Value
933	0.000000	0.000000	0.333333	Low Value
934	0.000000	0.666667	1.000000	High Value

935 rows × 4 columns

**Gambar 9.** Penempatan Label *Cluster* pada Data RFM Normalisasi

## 5.2 Pengukuran Keakuratan Hasil *Cluster*

Pengukuran kluster atau evaluasi hasil *clustering* sangat penting dilakukan agar dapat memastikan bahwa proses clustering yang telah dilakukan telah menghasilkan kluster yang baik dan sesuai dengan tujuan awal dari segmentasi data. Proses pengukuran hasil kluster dapat dilihat sebagai berikut.

a. Bangkitkan Nilai SSW (*Sum of Square Within cluster*)

$$C0 = \frac{0,614775858 + 0,615040805 + 0,615040805 + \dots}{409}$$

$$C1 = \frac{0,38840951 + 0,29214098 + 0,31155647 + \dots}{526}$$

**Tabel 17.** Nilai SSW

SSW	
SSW0	0,821368
SSW1	0,331035

b. Bangkitkan Nilai SSB

Selanjutnya akan dihitung *Sum of Squared Between (SSB)* 2 kluster antara kluster 0 (C0) ke kluster 1 (C1). Dapat dilihat pada sebagai berikut.

$$(C0, C1) SSW1 = \sqrt{(0,06960816 - 0,05593217)^2} + \dots$$

Hasil perhitungan *Sum of Squared Between (SSB)* antara kluster 0 (C0) dan kluster 1 (C1), serta antara kluster 2 (C2) didapatkan sebagai berikut:

**Tabel 18.** Nilai SSB

SSB	C0	C1
C0	0	0,543846
C1	0,543846	0

c. Mencari nilai Rasio

Untuk menentukan nilai *Rmax* pada hasil perhitungan SSB, diperlukan mencari nilai terbesar terlebih dahulu dari hasil perhitungan SSB yang telah dihitung sebelumnya yang mana nilai R Max adalah 0,543846. Setelah mendapatkan nilai *Rmax*, langkah selanjutnya adalah menghitung nilai rasio (*Ratio*) sebagai berikut.

$$Ratio = \frac{0,821368+0,331035}{0,543846} = 1,430059796$$

d. Nilai *DBI Score*

Dari hasil yang sudah didapatkan maka selanjutnya adalah menentukan nilai *DBI Score* yang mana dapat dilihat sebagai berikut.

$$Davies\ Bouldin\ Index = \frac{1,430059796}{2} = 0,715029898$$

Hasil evaluasi menggunakan *Davies Bouldin Index* menunjukkan nilai sebesar 0,715029898. Jika nilai *cluster* mendekati angka 0, maka *cluster* baik atau optimal.

### 5.3 Pembahasan

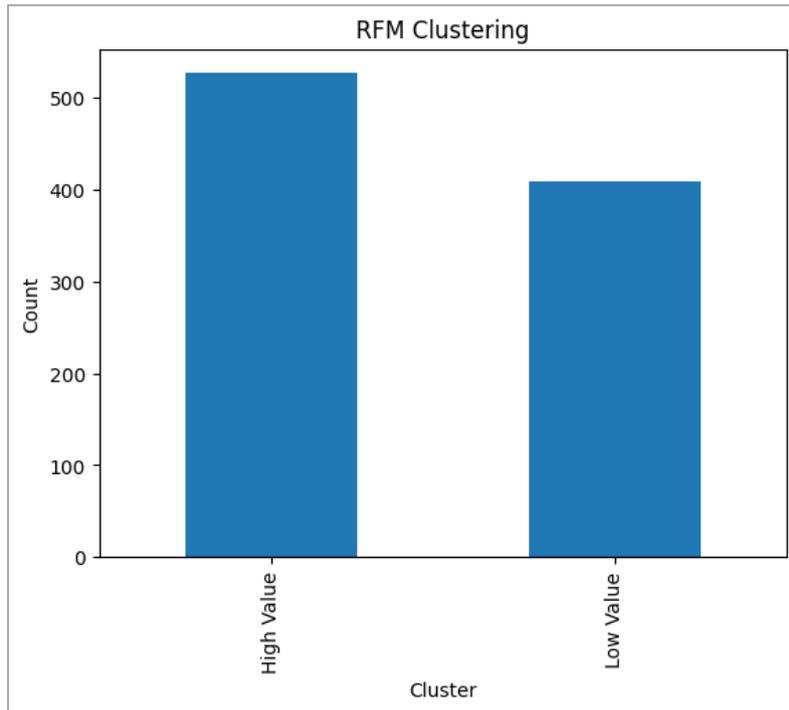
Dalam penelitian ini, data dari sebuah toko di Bogor bernama Hallogen digunakan untuk melakukan segmentasi pelanggan dengan menggunakan model RFM dan *Fuzzy C-Means*. Pemilik toko ingin mengetahui karakteristik pelanggan mereka secara lebih jelas. Data yang digunakan meliputi data selama 2 periode, yaitu dari September 2020 hingga September 2022. Lalu akan dilakukan pembuatan data pelanggan dengan menggunakan model RFM dengan variabel yang digunakan yaitu tanggal pembelian, total transaksi, dan total uang yang dikeluarkan oleh pelanggan.

Setelah itu, data di-*preprocessing* dan di transformasi seperti normalisasi lalu diaplikasikan pada metode *Fuzzy C-Means*. Metode ini memungkinkan pelanggan bergabung dalam lebih dari satu kelompok dengan tingkat keanggotaan yang berbeda-beda. Penelitian ini awalnya memberikan 3 kelompok pelanggan, namun setelah dilakukan evaluasi, pelanggan lebih cocok untuk di-*cluster* dalam 2 kelompok. Dalam kasus ini, pelanggan dibagi menjadi 2 kelompok, yaitu *High Value* dan *Low Value*. Dalam pengujian, nilai *dbi* yang didapatkan adalah 0,71, yang menunjukkan hasil *clustering* cukup baik atau optimal. Hasil klastering dan karakteristik kelompok pelanggan dapat membantu pemilik toko dalam memahami profil pelanggannya, seperti perilaku pembelian, preferensi produk, dan tingkat loyalitas, sehingga dapat dikembangkan strategi pemasaran yang lebih efektif dan meningkatkan retensi pelanggan.

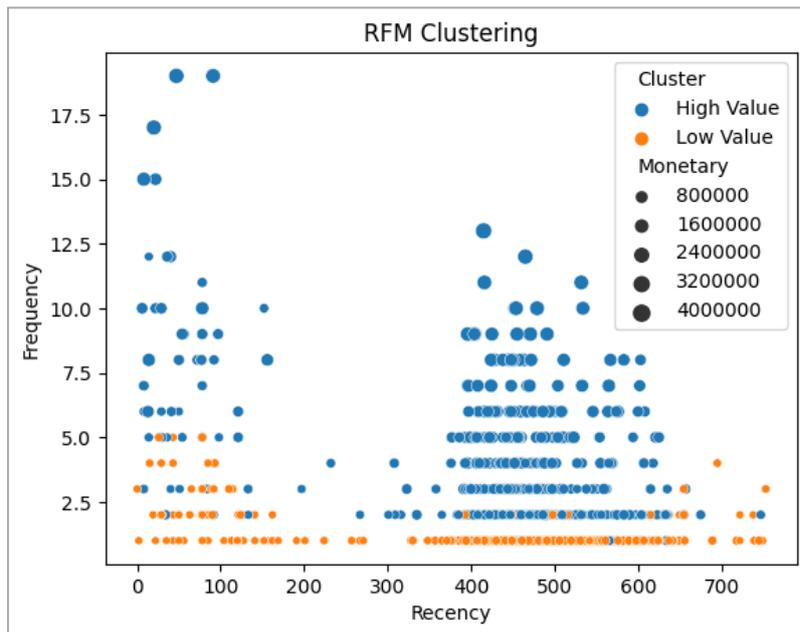
	<b>Id_Username</b>	<b>Recency</b>	<b>Frequency</b>	<b>Monetary</b>	<b>R</b>	<b>F</b>	<b>M</b>	<b>Cluster</b>
0	09035koc6e	14	12	372600	4	4	3	High Value
1	0albi_	22	10	1286600	4	4	4	High Value
2	17340_	430	2	648000	1	2	4	High Value
3	1904faris	444	2	898000	1	2	4	High Value
4	211209.ela	308	4	954000	1	3	4	High Value
...	...	...	...	...	...	...	...	...
930	zulfamarsalia	441	1	289000	1	1	2	Low Value
931	zulfiaswinn	141	1	31000	2	1	1	Low Value
932	zuroidadiana	449	1	349000	1	1	2	Low Value
933	zyadakebuli	461	1	349000	1	1	2	Low Value
934	zyahamida	444	3	873000	1	3	4	High Value

935 rows × 8 columns

**Gambar 10.** Hasil Pengelompokan Data dalam Kelompok-Kelompok (*Cluster*)



**Gambar 11.** Jumlah Data dalam Setiap Kelompok *Cluster*



**Gambar 12.** Hasil Pengelompokan Data dalam Bentuk Plot (Grafik)

Dari grafik yang telah ditunjukkan, dapat disimpulkan bahwa data pelanggan telah berhasil dipilah ke dalam 2 kelompok yang berbeda, yaitu kelompok *Low Value* dan kelompok *High Value*. Dari grafik tersebut terdapat jumlah data yang terdapat di masing-masing kelompok *cluster*. Kelompok *Low Value* memiliki jumlah data sebanyak 409, sedangkan kelompok *High Value* memiliki jumlah data sebanyak 526.

## BAB VI KESIMPULAN DAN SARAN

### 6.1 Kesimpulan

Penelitian ini menggunakan metode *Fuzzy C-Means* (FCM) untuk melakukan segmentasi pelanggan pada sebuah toko di Bogor berdasarkan variabel *Recency*, *Frequency*, dan *Monetary* (RFM). Skor RFM diberikan kepada pelanggan dengan rentang nilai dari 1 (terendah) hingga 4 (tertinggi), dan kemudian dinormalisasi menggunakan metode *min-max scaler*. Setelah melalui proses *preprocessing data* dan evaluasi jumlah *cluster*, ditemukan bahwa pelanggan dapat dibagi menjadi 2 kelompok yang disebut *High Value* dan *Low Value*.

*Indeks Davies-Bouldin* (DBI) adalah salah satu metrik yang digunakan untuk mengukur kualitas segmentasi pada metode *clustering*. Nilai DBI berkisar antara 0 hingga tak terbatas, dengan nilai yang lebih rendah menunjukkan segmentasi yang lebih baik. Dalam konteks penelitian ini, metode *Fuzzy C-Means* (FCM) mencapai nilai DBI sebesar 0,7.

*Fuzzy C-Means* merupakan metode yang efektif dalam segmentasi pelanggan. Dalam konteks segmentasi pelanggan ini, *Fuzzy C-Means* dapat memanfaatkan dengan baik informasi RFM yang telah dibentuk, sehingga menghasilkan pembagian kelompok yang relevan. *Fuzzy C-Means* juga sensitif terhadap perbedaan nilai RFM antara pelanggan, dan mampu menghasilkan kelompok yang fokus pada atribut tersebut. Dengan demikian, *Fuzzy C-Means* mampu memberikan pembagian kelompok yang lebih relevan dan signifikan dalam konteks penelitian ini, yang sangat penting dalam pengembangan strategi pemasaran yang efektif.

### 6.2 Saran

Saran pada penelitian ini adalah melakukan eksplorasi terhadap algoritma *clustering* lainnya, seperti *Hierarchical Clustering*, DBSCAN, atau *Gaussian Mixture Models*, untuk memperoleh pemahaman yang lebih lengkap tentang segmentasi pelanggan dalam konteks yang berbeda. Dalam proses *clustering*, terdapat beberapa hal yang perlu diperhatikan agar kualitas *clustering* dapat ditingkatkan. Pertama, pemilihan variabel yang digunakan dalam model RFM harus dipertimbangkan dengan matang, karena terdapat kemungkinan variabel lain yang lebih relevan dengan karakteristik pelanggan. Kedua, jumlah *cluster* optimal dapat bervariasi tergantung pada data dan bisnis yang dijalankan, sehingga disarankan untuk melakukan evaluasi jumlah *cluster* yang lebih teliti dan menguji beberapa pilihan jumlah *cluster*. Terakhir, selain nilai dbi, terdapat beberapa metode evaluasi *clustering* lainnya seperti *Silhouette Score*, *CH Index*, dan *Dunn Index* yang dapat digunakan untuk mengevaluasi kualitas *clustering* secara lebih terperinci. Dengan memperhatikan faktor-faktor tersebut, proses *clustering* dapat menghasilkan segmentasi pelanggan yang lebih baik dan bermanfaat bagi bisnis atau toko yang dijalankan.

## DAFTAR PUSTAKA

- Adiana, B. E., Soesanti, I., & Permanasari, A. E. (2018). Analisis Segmentasi Pelanggan Menggunakan Kombinasi Rfm Model Dan Teknik *Clustering*. *Jurnal Terapan Teknologi Informasi* <https://doi.org/10.21460/jutei.2018.21.76>, 2(1), 23–32.
- Annisa, K., Serasi Ginting, B., & Syari, M. A. (n.d.). Penerapan *Data Mining* Pengelompokan Data Pengguna Air Bersih Berdasarkan Keluhannya Menggunakan Metode *Clustering* Pada PDAM Langkat.
- Basri, B., Gata, W., & Risnandar, R. (2020). Analisis Loyalitas Pelanggan Berbasis *Model Recency, Frequency, dan Monetary (RFM)* dan *Decision Tree* pada PT. Solo. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 7(5), 943. <https://doi.org/10.25126/jtiik.2020752284>
- Boentarman, M., Rostianingsih, S., & Setiawan, A. (n.d.). Penerapan Segmentasi Pelanggan dengan Menggunakan Metode *K-Means Clustering* Pada Sistem *Customer Relationship Management* di PT. Titess.
- Dista, T. M., & Abdulloh, F. F. (2022). *Clustering* Pengunjung *Mall* Menggunakan Metode *K-Means* dan *Particle Swarm Optimization*. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 6(3), 1339. <https://doi.org/10.30865/mib.v6i3.4172>
- Gie, W., & Jollyta, D. (2020). Perbandingan *Euclidean* dan *Manhattan* Untuk Optimasi *Cluster* Menggunakan *Davies Bouldin Index*: Status Covid-19 Wilayah Riau. *Prosiding Seminar Nasional Riset Dan Information Science (SENARIS)*, 2, 187–191.
- Harani, N. H., Prianto, C., & Nugraha, F. A. (n.d.). Segmentasi Pelanggan Produk Digital Service Indihome Menggunakan Algoritma K-Means Berbasis Python. *Jurnal Manajemen Informatika (JAMIKA)*. <https://doi.org/10.34010/jamika.v10i2>.
- Ilham, A., Setiawan, N. Y., & Afirianto, T. (2020). Analisis Segmentasi Pelanggan Kartu Prabayar Kabupaten Malang dengan RFM Model Menggunakan Metode *Fuzzy C-Means Clustering* (Studi Kasus : PT. XYZ). 4(8), 2487–2498. <http://j-ptiik.ub.ac.id>
- Monalisa, S., Bahari, P., Fitra, D., Uin, K., Riau, S., Soebrantas, H. R., & 15, K. M. (n.d.). *Analisis Segmentasi dalam Penentuan Target Pemasaran Mahasiswa UIN Suska Riau dengan Metode Fuzzy C Means*.
- Pramudiansyah, A. P. (2021). Segmentasi Pelanggan Menggunakan Algoritma *K-Means* Berdasarkan *Model Recency Frequency Monetary*. *Jurnal Ilmiah Ilmu Komputer*, 7(2), 06–19. <https://doi.org/10.35329/jiik.v7i2.201>

- Prasetyo, S. S., Mustafid, M., & Hakim, A. R. (2020). Penerapan *Fuzzy C-Means* Kluster Untuk Segmentasi Pelanggan *E-Commerce* Dengan Metode *Recency Frequency Monetary (Rfm)*. *Jurnal Gaussian*, 9(4), 421–433. <https://doi.org/10.14710/j.gauss.v9i4.29445>
- Rahakbauw, D. L., Ilwaru, V. Y. I., & Hahury, M. H. (2017). Implementasi *Fuzzy C-Means Clustering* Dalam *Implementation Of Fuzzy C-Means Clustering In*. *Jurnal Ilmu Matematika Dan Terapan*, 11, 1–12.
- Rukmana, S. H. (2020). PENINGKATAN JUMLAH NASABAH di BRI SYARIAH KCP PARE KEDIRI. 86.
- Septiani, I. W., Fauzan, Abd. C., & Huda, M. M. (2022). Implementasi Algoritma *K-Medoids* Dengan Evaluasi *Davies-Bouldin-Index* Untuk Klasterisasi Harapan Hidup Pasca Operasi Pada Pasien Penderita Kanker Paru-Paru. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 3(4), 556. <https://doi.org/10.30865/json.v3i4.4055>
- Setiawan, Amani, H., & Tripiawan, W. (2021). Perancangan Segmentasi Pelanggan Dengan Metode *Clustering K-Means* Dan Model Rfm Pada Klinik Kecantikan *Seoul Secret Design of Customer Segmentation With Clustering K-Means Method and Rfm Model in Seoul Secret Beauty Clinic*. *E-Proceeding of Engineering*, 8(2), 2286–2293.
- Taqwim, & Dkk. (2019). Analisis Segmentasi Pelanggan Dengan RFM Model Pada Pt. Arthamas Citra Mandiri Menggunakan Metode *Fuzzy C-Means Clustering*. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(2), 1986–1993.
- Thamrin, N., & Wijayanto, A. W. (2021). *Comparison of Soft and Hard Clustering: A Case Study on Welfare Level in Cities on Java Island*. *Indonesian Journal of Statistics and Its Applications*, 5(1), 141–160. <https://doi.org/10.29244/ijsa.v5i1p141-160>
- Vol, M. I. (2022). *Media Informatika* Vol. 21 No. 2 (2022) 113. 21(2), 113–123.
- Wardani, N. W., Arnidya, D. J., Agus, I. N., Putra, S., Made, N., & Rosa, M. (2022). Prediksi Pelanggan Loyal Menggunakan Metode *Naïve Bayes* Berdasarkan Segmentasi Pelanggan dengan Pemodelan RFM. 12(2), 113–124.
- Yarif, N. R., & WindartoS. (2018). Implementasi Algoritma *Fuzzy C-Means* Dan Metode *Recency Frequency Monetary ( Rfm )* Pada Aplikasi Data. *Sebatik*, 88–94.

# **LAMPIRAN**

## Lampiran 1. Surat Keterangan Penelitian



YAYASAN PAKUAN SILIWANGI  
**Universitas Pakuan**  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
*Unggul, Mandiri & Berkarakter Dalam Bidang MIPA*

**KEPUTUSAN DEKAN**  
**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**  
**UNIVERSITAS PAKUAN**  
**No. : 1342/KEP/D/FMIPA/IV/2023**

**T E N T A N G**

**PENGANGKATAN PEMBIMBING TUGAS AKHIR**  
**PADA PROGRAM STUDI ILMU KOMPUTER**  
**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**  
**UNIVERSITAS PAKUAN**

**DEKAN FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**  
**UNIVERSITAS PAKUAN,**

- Menimbang : a. bahwa setiap mahasiswa tingkat akhir Program Strata Satu (S1) harus melaksanakan Tugas Akhir sebagaimana tercantum di dalam kurikulum setiap Program Studi di lingkungan Fakultas MIPA Universitas Pakuan.  
b. bahwa untuk pelaksanaan Tugas Akhir diperlukan pengawasan dari pembimbing.  
c. bahwa sehubungan dengan point a dan b di atas perlu dituangkan dalam suatu Keputusan Dekan.
- Mengingat : 1. Undang-undang RI No.: 20 Tahun 2003 tentang Sistem Pendidikan Nasional.  
2. Peraturan Pemerintah No.: 60 Tahun 1999 tentang Pendidikan Tinggi.  
3. Statuta Universitas Pakuan Tahun 2019.  
4. Surat Keputusan Rektor Nomor: 35/KEP/REK/VIII/2020 tanggal 03 Agustus 2020 tentang Pemberhentian Dekan dan Wakil Dekan Masa Bakti 2015-2020 serta Pengangkatan Dekan dan Wakil Dekan Masa Bakti 2020-2025 di lingkungan Universitas Pakuan.  
5. Ketentuan Akademik yang tercantum dalam Buku Panduan Studi Fakultas MIPA, Universitas Pakuan Tahun 2022.

Memperhatikan : Usulan dari Ketua Program Studi Ilmu Komputer FMIPA UNPAK.

**M E M U T U S K A N**

- Menetapkan :  
Pertama : Mengangkat pembimbing yang namanya tersebut di bawah ini :  
1. Pembimbing Utama : Asep Denih, S.Kom., M.Sc., PhD  
2. Pembimbing Pendamping : Irma Anggraeni, M.Kom.

Untuk membimbing dalam rangka melaksanakan tugas akhir bagi mahasiswa :  
Nama : Yusuf Fadilah Rukmana  
NPM : 065118220  
Program Studi : Ilmu Komputer  
Judul Skripsi : Analisis Segmentasi Pelanggan Menggunakan Algoritma Fuzzy C-Means dan Model Recency, Frequency, Monetary

- Kedua : Kepada para pembimbing diharapkan dapat menjalankan tugasnya sebagai pembimbing dengan sebaik-baiknya.
- Ketiga : Dalam waktu 1 (satu) bulan setelah diterbitkannya SK ini, mahasiswa wajib melaksanakan Seminar Rencana Penelitian yang diselenggarakan oleh Program Studi Ilmu Komputer dengan dihadiri oleh Pembimbing dan Penguji.
- Keempat : Dana untuk honorarium pembimbing dibebankan kepada mahasiswa yang ketentuannya diatur oleh Fakultas MIPA.
- Kelima : Surat Keputusan ini berlaku untuk jangka waktu 1 (satu) tahun sejak tanggal ditetapkan sampai dengan mahasiswa tersebut Lulus Sidang/Ujian Skripsi, dengan ketentuan akan diadakan perubahan/perbaikan sebagaimana mestinya bila dikemudian hari terdapat kekeliruan dalam penetapannya.

Ditetapkan di : Bogor  
Pada tanggal : 05 April 2023

Dekan,



Asep Denih, S.Kom., M.Sc., Ph.D.

Tembusan :

1. Yth. Ketua Program Studi Ilmu Komputer;
2. Yth. Asep Denih, S.Kom., M.Sc.,PhD;
3. Yth. Irma Anggraeni, M.Kom.;
4. Arsip.

## Lampiran 2. Surat Perizinan Perusahaan



### PEMERINTAH REPUBLIK INDONESIA

#### PERIZINAN BERUSAHA BERBASIS RISIKO NOMOR INDUK BERUSAHA: 0405230061705

Berdasarkan Undang-Undang Nomor 6 Tahun 2023 tentang Penetapan Peraturan Pemerintah Pengganti Undang-Undang Nomor 2 Tahun 2022 tentang Cipta Kerja Menjadi Undang-Undang, Pemerintah Republik Indonesia menerbitkan Nomor Induk Berusaha (NIB) kepada:

- |  |   |
|--|---|
| 1. Nama Pelaku Usaha                                     | : ACHMAD TAMIMI   |
| 2. Alamat  | : BOJONG, Desa/Kelurahan Bojong Pondok Terong, Kec. Cipayung, Kota Depok, Provinsi Jawa Barat |
| 3. Nomor Telepon Seluler                                 | : +628988031713   |
| Email  | : -   |
| 4. Kode Klasifikasi Baku Lapangan Usaha Indonesia (KBLI) | : Lihat Lampiran  |
| 5. Skala Usaha   | : Usaha Mikro   |

NIB ini berlaku di seluruh wilayah Republik Indonesia selama menjalankan kegiatan usaha dan berlaku sebagai hak akses kepebeanaan, pendaftaran kepesertaan jaminan sosial kesehatan dan jaminan sosial ketenagakerjaan, serta bukti pemenuhan laporan pertama Wajib Laporkan Ketenagakerjaan di Perusahaan (WLKP).

Pelaku Usaha dengan NIB tersebut di atas dapat melaksanakan kegiatan berusaha sebagaimana terlampir dengan tetap memperhatikan ketentuan peraturan perundang-undangan.

NIB ini merupakan perizinan tunggal yang berlaku sebagai sertifikasi jaminan produk halal berdasarkan pernyataan mandiri pelaku usaha dan setelah memperoleh pembinaan dan/atau pendampingan Proses Produk Halal (PPH) dari Pemerintah Pusat dan Pemerintah Daerah sesuai dengan ketentuan peraturan perundang-undangan.

Diterbitkan di Jakarta, tanggal: 4 Mei 2023

**Menteri Investasi/  
Kepala Badan Koordinasi Penanaman Modal,**



Ditandatangani secara elektronik

Dicetak tanggal: 4 Mei 2023

1. Dokumen ini diterbitkan sistem OSS berdasarkan data dari Pelaku Usaha, tersimpan dalam sistem OSS, yang menjadi tanggung jawab Pelaku Usaha.
2. Dalam hal terjadi kekeliruan isi dokumen ini akan dilakukan perbaikan sebagaimana mestinya.
3. Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan oleh BSR-BSSN.
4. Data lengkap Perizinan Berusaha dapat diperoleh melalui sistem OSS menggunakan hak akses.





**PEMERINTAH REPUBLIK INDONESIA**  
**PERIZINAN BERUSAHA BERBASIS RISIKO**  
**LAMPIRAN**  
**NOMOR INDUK BERUSAHA: 0405230061705**

Lampiran berikut ini memuat daftar bidang usaha untuk:

No.	Kode KBLI	Judul KBLI	Lokasi Usaha	Tingkat Risiko	Perizinan Berusaha		
					Jenis	Status	Keterangan
1	45406	Perdagangan Eceran Suku Cadang Sepeda Motor Dan Aksesorinya	Deparis Residence B11/7, Desa/Kelurahan Sasak Panjang, Kec. Tajurhalang, Kab. Bogor, Provinsi Jawa Barat Kode Pos: 16518	Rendah	NIB	Terbit	-
2	47854	Perdagangan Eceran Kaki Lima Dan Los Pasar Tas, Dompot, Koper, Ransel Dan Sejenisnya	BOJONG RT005 RW 002 BOJONG PONDOK TERONG, RT/RW 05/02, BOJONG PONDOK TERONG, CIPAYUNG, Desa/Kelurahan Bojong Pondok Terong, Kec. Cipayung, Kota Depok, Provinsi Jawa Barat Kode Pos: 16444	Rendah	NIB	Terbit	-

1. Dengan ketentuan bahwa NIB tersebut hanya berlaku untuk Kode dan Judul KBLI yang tercantum dalam lampiran ini.
2. Pelaku Usaha wajib memenuhi persyaratan dan/atau kewajiban sesuai Norma, Standar, Prosedur, dan Kriteria (NSPK) Kementerian/Lembaga (K/L).
3. Pengawasan pemenuhan persyaratan dan/atau kewajiban Pelaku Usaha dilakukan oleh Kementerian/Lembaga/Pemerintah Daerah terkait.
4. Lampiran ini merupakan bagian tidak terpisahkan dari dokumen NIB tersebut.

1. Dokumen ini diterbitkan sistem OSS berdasarkan data dari Pelaku Usaha, tersimpan dalam sistem OSS, yang menjadi tanggung jawab Pelaku Usaha.
2. Dalam hal terjadi kekeliruan isi dokumen ini akan dilakukan perbaikan sebagaimana mestinya.
3. Dokumen ini telah ditandatangani secara elektronik menggunakan sertifikat elektronik yang diterbitkan oleh BSR-E-BSSN.
4. Data lengkap Perizinan Berusaha dapat diperoleh melalui sistem OSS menggunakan hak akses.



### Lampiran 3. Data Penjualan Toko Hallogen

	No	No. Pesanan	Id_Username	Tanggal	Total_Item	Sales	Harga_Produk	Total_Pembayaran
0	1	200901M3FQKR09	dezsyf	2020-09-01 00:00:00	1	135,000	135,000	135,000
1	2	200901J4GQ0GS6	dezsyf	2020-09-02 00:00:00	3	135,000	135,000	405,000
2	3	200902B33JKAME	sausanbatarfie	2020-09-02 00:00:00	1	29,000	29,000	29,000
3	4	200907M3390NKQ	sausanbatarfie	2020-09-03 00:00:00	2	29,000	29,000	58,000
4	5	200902N26UFBJB	sausanbatarfie	2020-09-05 00:00:00	4	29,000	29,000	116,000
5	6	20090994TTWDNA	dessyantana00	2020-09-09 00:00:00	1	99,000	99,000	99,000
6	7	2009099B5RCH9D	teger.ayu	2020-09-09 00:00:00	1	399,000	399,000	399,000
7	8	200909ABARQGXD	larasatiart_23	2020-09-09 00:00:00	1	199,000	199,000	199,000
8	9	200910BR80S9J2	pustakahaura	2020-09-10 00:00:00	1	99,000	99,000	99,000
9	10	200911E2G2AY84	teger.ayu	2020-09-11 00:00:00	1	135,000	135,000	135,000
10	11	200913MCC6CBUQ	sekarbeauty.id	2020-09-13 00:00:00	1	99,000	99,000	99,000
11	12	200914MQAHE3T6	virdaazmin	2020-09-13 00:00:00	1	325,000	325,000	325,000
12	13	2009194UMQXE8T	mininecha	2020-09-19 00:00:00	1	99,000	99,000	99,000
13	14	2009195BEDRTF4	nayla.w.p	2020-09-19 00:00:00	1	150,000	150,000	150,000
14	15	2009203E81DKQE	ismaaaaa0	2020-09-20 00:00:00	1	29,000	29,000	29,000
15	16	2009208R75AHNQ	ismaaaaa0	2020-09-20 00:00:00	1	29,000	29,000	29,000
16	17	200929VW2OR9D	devita0.0	2020-09-29 00:00:00	1	29,000	29,000	29,000
17	18	2009292N5B6JG3	devita0.0	2020-09-29 00:00:00	1	29,000	29,000	29,000
18	19	2009293HHQ29FG	devita0.0	2020-09-29 00:00:00	1	29,000	29,000	29,000
19	20	20100133HJLW23	pustakahaura	2020-10-01 00:00:00	1	29,000	29,000	29,000
20	21	20100166V7B4UF	pustakahaura	2020-10-01 00:00:00	1	29,000	29,000	29,000

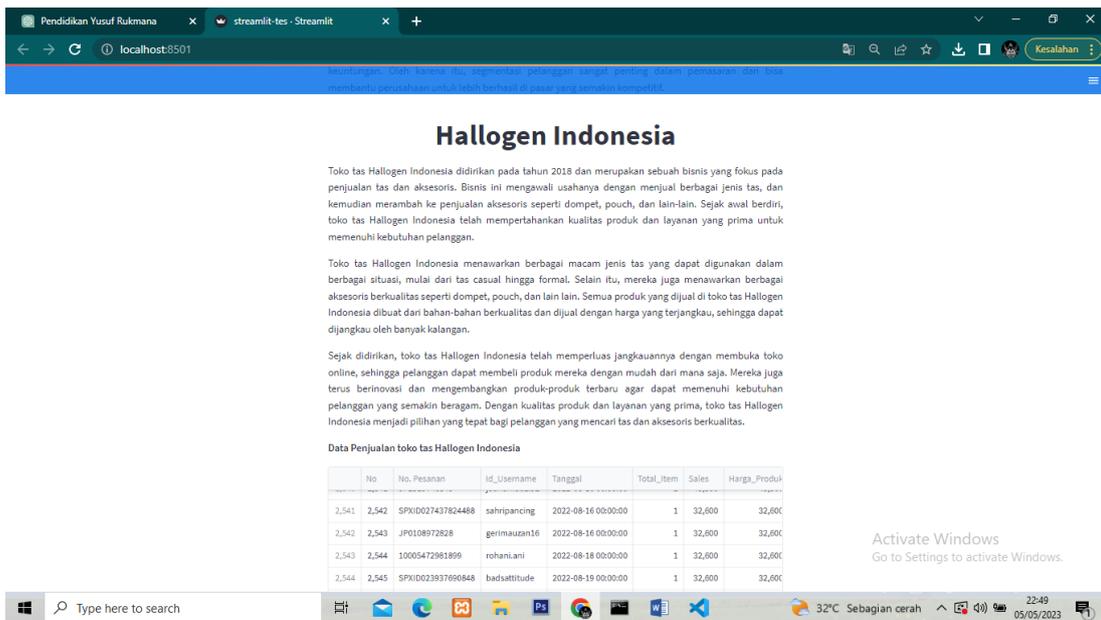
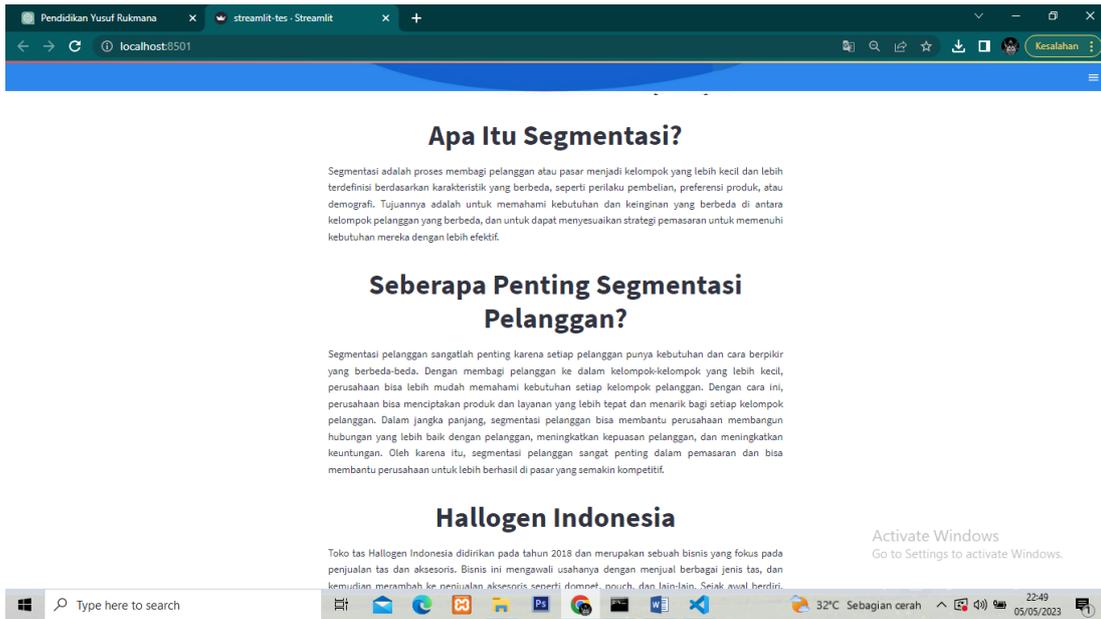
	No	No. Pesanan	Id_Username	Tanggal	Total_Item	Sales	Harga_Produk	Total_Pembayaran
2,576	2,577	10005636556015	yudotujuh	2022-09-06 00:00:00	1	45,500	45,500	45,500
2,577	2,578	SPXID027204169739	mohammadyuk	2022-09-06 00:00:00	1	32,600	32,600	32,600
2,578	2,579	SPXID025131329539	0albi_	2022-09-06 00:00:00	1	32,600	32,600	32,600
2,579	2,580	JP3605402052	dhilahshop.plg	2022-09-08 00:00:00	1	32,600	32,600	32,600
2,580	2,581	SPXID029092744749	syfakhoerunisa_	2022-09-09 00:00:00	1	32,600	32,600	32,600
2,581	2,582	4082799808	nyomanutami	2022-09-09 00:00:00	5	32,600	32,600	163,000
2,582	2,583	SPXID022497633629	09035koc6e	2022-09-12 00:00:00	1	32,600	32,600	32,600
2,583	2,584	10005681507613	09035koc6e	2022-09-14 00:00:00	1	32,600	32,600	32,600
2,584	2,585	10005710427734	n211183r	2022-09-14 00:00:00	1	54,000	54,000	54,000
2,585	2,586	JP9510750216	mauraadibah	2022-09-14 00:00:00	1	32,600	32,600	32,600
2,586	2,587	SPXID021716222399	foloe	2022-09-14 00:00:00	1	32,600	32,600	32,600
2,587	2,588	GK-11-631869493	adrian1302	2022-09-13 00:00:00	1	32,600	32,600	32,600
2,588	2,589	SPXID020024826149	4412r13c7l	2022-09-14 00:00:00	1	32,600	32,600	32,600
2,589	2,590	JP0265222414	nyomanutami	2022-09-15 00:00:00	1	99,000	99,000	99,000
2,590	2,591	SPXID023198575389	alrafa_shop	2022-09-20 00:00:00	1	32,500	32,500	32,500
2,591	2,592	SPXID029843738689	mohammadyuk	2022-09-20 00:00:00	1	32,500	32,500	32,500
2,592	2,593	10005773172336	dzaki1899	2022-09-20 00:00:00	1	32,500	32,500	32,500
2,593	2,594	SPXID022292938039	nila08	2022-09-20 00:00:00	1	32,500	32,500	32,500
2,594	2,595	SPXID020695093779	mauraadibah	2022-09-22 00:00:00	1	32,500	32,500	32,500
2,595	2,596	SD-2-0HY7CRH3H1B4	diskyantarez	2022-09-26 00:00:00	1	77,500	77,500	77,500
2,596	2,597	SPXID029297523939	foloe	2022-09-28 00:00:00	1	32,500	32,500	32,500

#### Lampiran 4. Visualisasi Website

**Analisis Segmentasi Pelanggan Menggunakan Algoritma Fuzzy C-Means Dan Model (RFM) Recency, Frequency, Monetary.**

**Apa Itu Segmentasi?**

Segmentasi adalah proses membagi pelanggan atau pasar menjadi kelompok yang lebih kecil dan lebih terdefinisi berdasarkan karakteristik yang berbeda, seperti perilaku pembelian, preferensi produk, atau demografi. Tujuan adalah untuk memahami lebih baik kebutuhan pelanggan yang berbeda, di antara



Pendidikan Yusuf Rukmana x streamlit-tes - Streamlit x +

localhost:8501

perusahaan membuat strategi pemasaran dan bisnis yang lebih efektif dan efisien.

## Model Recency, Frequency, dan Monetary (RFM)

### Recency

Recency dalam RFM model adalah faktor yang mengukur seberapa baru terakhir kali pelanggan melakukan transaksi atau interaksi dengan bisnis. Dalam RFM model, Recency biasanya diukur dalam jumlah hari atau bulan antara transaksi terakhir pelanggan dan saat ini. Semakin kecil nilai Recency, artinya semakin baru terakhir kali pelanggan berinteraksi dengan bisnis, dan semakin besar kemungkinan bahwa pelanggan tersebut masih aktif dan berniat untuk melakukan transaksi lebih lanjut. Berikut contoh data hasil olahan RFM model untuk Recency:

Data Recency

	id_username	LastPurchaseDate	Recency
0	0903kocbe	2022-09-14 00:00:00	14
1	0a8bl_	2022-09-06 00:00:00	22
2	17340_	2021-07-25 00:00:00	430
3	1904faris	2021-07-11 00:00:00	444
4	211209.ela	2021-11-24 00:00:00	308
5	441213c7i	2022-09-14 00:00:00	14
6	57jklahyov	2021-06-10 00:00:00	475
7	___yarihdy	2022-05-24 00:00:00	127
8	_bopung.id	2021-05-15 00:00:00	501

Activate Windows  
Go to Settings to activate Windows.

Type here to search

32°C Sebagian cerah 22:49 05/05/2023

Pendidikan Yusuf Rukmana x streamlit-tes - Streamlit x +

localhost:8501

## Clustering dengan Fuzzy C-Means

Sebelum dilakukan proses clustering menggunakan metode Fuzzy C-means pada nilai RFM score, sebaiknya nilai RFM score dinormalisasi terlebih dahulu. Hal ini dilakukan untuk menghindari terjadinya bias dalam proses clustering dan memastikan bahwa setiap faktor RFM mempunyai bobot yang sama. Normalisasi bertujuan untuk mengubah setiap faktor RFM menjadi skala yang sama sehingga memiliki bobot yang sama dalam proses clustering. Ada beberapa metode normalisasi yang dapat digunakan, seperti Min-Max Normalization, Z-Score Normalization, dan Decimal Scaling Normalization dan pada data ini diproses dengan menggunakan Min-Max Normalization.

Data RFM Score yang dinormalisasi

(9x3, 3)

	R	F	M
0	1	1	0.6667
1	1	1	1
2	0	0.3333	1
3	0	0.3333	1
4	0	0.6667	1
5	1	0.6667	0.6667
6	0	0.3333	1
7	0.3333	0.3333	0
8	0	0.3333	1
9	0.6667	0.6667	0

Activate Windows  
Go to Settings to activate Windows.

Type here to search

32°C Sebagian cerah 22:50 05/05/2023



### Label Tiap Cluster

Pelabelan cluster adalah proses memberikan kategori pada setiap kelompok hasil clustering pada data pelanggan. Setelah mendapatkan probabilitas setiap data termasuk ke dalam kelompok tertentu, maka kita dapat menentukan kelompok mana yang cocok untuk setiap data dengan memilih nilai probabilitas tertinggi. Dengan mengetahui kelompok-kelompok yang ada, kita dapat memahami karakteristik dan pola yang terdapat pada setiap kelompok, yang dapat digunakan untuk mengambil keputusan bisnis yang lebih baik.

Data Pelanggan yang sudah dilabeli tiap clusternya

	id_Username	Recency	Frequency	Monetary	R	F	M	Cluster
0	09035kocfe	14	12	372,600	4	4	3	High Value
1	0a0bl_	22	10	1,286,600	4	4	4	High Value
2	17340_	430	2	648,000	1	2	4	High Value
3	1904faris	444	2	898,000	1	2	4	High Value
4	211209.ela	308	4	954,000	1	3	4	High Value
5	4412x13v7l	14	5	492,600	4	3	3	High Value
6	57kiahyov	475	2	548,000	1	2	4	High Value
7	___yairhdy	127	2	62,000	2	2	1	Low Value
8	_bojung.id	501	2	798,000	1	2	4	High Value
9	_refaariani	50	3	99,000	3	3	1	Low Value

Activate Windows  
Go to Settings to activate Windows.

### RFM Clustering

