

**KLASTERISASI LOYALITAS PELANGGAN BERDASARKAN  
ANALISIS RFM MENGGUNAKAN *GAUSSIAN  
MIXTURE MODEL (GMM)***

**SKRIPSI**

**OLEH:**

**YENI SOFHI MELIANI**

**218160008**



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MEDAN AREA  
MEDAN  
2025**

**UNIVERSITAS MEDAN AREA**

© Hak Cipta Di Lindungi Undang-Undang

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Document Accepted 1/8/25

Access From (repository.uma.ac.id)1/8/25

**KLASTERISASI LOYALITAS PELANGGAN BERDASARKAN  
ANALISIS RFM MENGGUNAKAN *GAUSSIAN  
MIXTURE MODEL (GMM)***

**SKRIPSI**

**Diajukan sebagai Salah Satu Syarat untuk Memperoleh  
Gelara Sarjana di Fakultas Teknik  
Universitas Medan Area**

**Oleh:**

**YENI SOFHI MELIANI**

**218160008**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS MEDAN AREA  
MEDAN  
2025**

**UNIVERSITAS MEDAN AREA**

© Hak Cipta Di Lindungi Undang-Undang

Document Accepted 1/8/25

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

Access From (repository.uma.ac.id)1/8/25

## HALAMAN PENGESAHAN

Judul Skripsi : Klasterisasi Loyalitas Pelanggan Berdasarkan Analisis RFM Menggunakan Gaussian Mixture Model (GMM)

Nama : Yeni Sofhi Meliani  
NPM : 218160008  
Fakultas : Teknik

Disetujui Oleh  
Komisi Pembimbing

  
Andre Hasudungan Lubis, S.Ti, M.Sc.  
Pembimbing



  
Andre Hasudungan Lubis, S.Ti, M.Sc.  
Pembimbing



  
Yeni Sofhi Meliani, M.Kom  
Teknik Informatika

Tanggal Lulus: 06 Maret 2025

## HALAMAN PERNYATAAN

Saya menyatakan bahwa skripsi yang saya susun, sebagai syarat memperoleh gelar sarjana merupakan hasil karya tulis saya sendiri. Adapun bagian-bagian tertentu dalam penulisan skripsi ini yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas sesuai dengan norma, kaidah, dan etika penulisan ilmiah.

Saya bersedia menerima sanksi pencabutan gelar akademik yang saya peroleh dan sanksi-sanksi lainnya dengan peraturan yang berlaku, apabila di kemudian hari ditemukan adanya plagiat dalam skripsi ini.

Medan, 06 Maret 2025



Yeni Sofhi Meliani  
218160008

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI  
TUGAS AKHIR/SKRIPSI/TESIS UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Medan Area, saya yang bertanda tangan di bawah ini:

Nama : Yeni Sofhi Meliani  
NPM : 218160008  
Program Studi : Teknik Informatika  
Fakultas : Teknik  
Jenis Karya : Skripsi

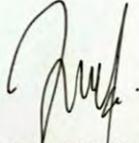
Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Medan Area **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul :

**“Klasterisasi Loyalitas Pelanggan Berdasarkan Analisis RFM Menggunakan *Gaussian Mixture Model* (GMM)”**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Medan Area berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan Skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Medan  
Pada tanggal : 06 Maret 2025  
Yang menyatakan

  
Yeni Sofhi Meliani  
218160008

## RIWAYAT HIDUP

Penulis bernama Yeni Sofhi Meliani lahir di RantauPrapat pada tanggal 19 Mei 2003 Penulis merupakan anak ke-1 dari 2 bersaudara dari pasangan Bapak Rudi Parmono dan Ibu Fitria

Penulis memulai pendidikan formalnya di:

- SD NEGERI 015863 Sei Renggas dan lulus pada tahun 2015
- kemudian melanjutkan ke SMP Negri 1 Rantau Utara lulus tahun 2018
- dan melanjutkan ke SMK Prayatna 2 Medan, lulus pada tahun 2021

Pada tahun 2021 penulis diterima sebagai mahasiswa Program Studi Teknik Informatika di Fakultas Teknik Universitas Medan Area

Selama menjalani masa perkuliahan, penulis aktif dalam berbagai kegiatan akademik maupun non-akademik, termasuk organisasi kemahasiswaan dan kegiatan sosial. Pengalaman ini menjadi bekal berharga dalam pengembangan diri, baik secara intelektual maupun emosional.

Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana pada Program Studi Teknik Informatika, dengan judul: “Klasterisasi Loyalitas Pelanggan Berdasarkan Analisis RFM Menggunakan *Gaussian Mixture Model* (GMM)”

Penulis berharap bahwa ilmu dan pengalaman yang diperoleh selama masa studi dapat menjadi landasan yang kuat dalam meniti karier dan memberikan kontribusi positif di tengah masyarakat.

## KATA PENGANTAR

Puji dan Syukur penulis panjatkan ke hadirat Allah SWT atas segala kelimpahan rahmat, taufik, dan hidayah-nya sehingga penulis dapat menyelesaikan skripsi ini yang berjudul “Klasterisasi Loyalitas Pelanggan Berdasarkan Analisa RFM Menggunakan Gaussian Mixture Model (GMM)”. Karya ini merupakan salah satu syarat untuk memperoleh gelar sarjana pada program studi Teknik Informatika Universitas Medan Area.

Skripsi ini bukan hanya hasil dari kerja keras secara akademik, tetapi juga refleksi dari perjalanan pribadi yang penuh dinamika. Setiap bab yang tertulis adalah hasil dari rasa ingin tahu penulis yang terus menyala. Penulis berharap skripsi ini tidak hanya menjadi dokumen formal, tetapi juga dapat menjadi inspirasi dan juga manfaat bagi pembaca, peneliti selanjutnya dan masyarakat luas.

Dalam penyusunan skripsi ini, penulis mendapatkan dukungan, motivasi dan bantuan dari banyak pihak. Oleh karena itu, dengan segala kerendahan hati, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. Dadan Ramdan, M.Eng., M.Sc., selaku Rektor Universitas Medan Area.
2. Bapak Dr. Eng. Supriatno, S.T., M.T., selaku Dekan Fakultas Teknik Universitas Medan Area.
3. Bapak Rizki Muliono, S.Kom., M.Kom., selaku Kepala Program Studi Teknik informatika.
4. Kepada bapak Andre Hasudungan Lubis, terimah kasih yang sebesar-besarnya atas segala bimbingan, kesabaran, dan ketulusan yang telah

bapak berikan. Di tengah kesubukan dan tanggung jawab yang begitu banyak bapak tetap meluangkan waktu untuk membimbing saya, memberi arahan yang penuh dengan makna, seta menyemangati saya. Tidak hanya ilmu yang saya dapatkan, tetapi juga pembelajaran hidup tentang ketekunan, integritas, dan keikhlasan dalam mendampingi mahasiswa. Bapak bukan hanya pembimbing akademik, tetapi juga panutan dalam bersikap dan berpikir. Terima kasih telah menjadi bagian penting dalam proses tumbuh dan belajar saya.

5. Terima kasih yang tidak akan pernah cukup diucapkan kepada Bapak Rudi Parmono sebagai cinta pertama penulis yaitu, Papa tercinta, sosok yang diam-diam selalu menjadi tiang paling kokoh dalam hidup penulis. Dalam setiap langkah yang penulis tempuh, ada doa yang tidak terdengar namun selalu terasa, ada kerja keras yang tidak pernah ditunjukkan namun jelas terlihat dari lelahnya raut wajah Papa setiap hari. Papa mungkin tidak banyak bicara, tapi melalui caranya menjaga, berkorban, dan terus mendorong anaknya untuk berdiri tegak, penulis belajar tentang arti tanggung jawab dan ketulusan. Terima kasih telah menjadi teladan tentang bagaimana menjadi kuat tanpa harus bersuara lantang, dan bagaimana mencintai tanpa harus banyak kata. Skripsi ini mungkin hanya setumpuk kertas bagi sebagian orang, tapi bagi penulis, ini adalah bukti kecil dari doa dan keringat Papa yang selama ini tidak pernah berhenti mengalir. Semoga karya ini bisa menjadi wujud balas budi, meski penulis tahu, tidak akan pernah sepadan dengan semua yang telah Papa berikan.

6. Terima kasih yang sedalam-dalamnya penulis sampaikan kepada Ibu Fitria sebagai sosok paling istimewa dalam hidup penulis, Mama tercinta. Untuk setiap doa yang tidak pernah putus, bahkan di saat penulis sendiri mulai ragu pada langkah yang diambil. Untuk setiap air mata yang jatuh diam-diam, setiap pelukan hangat yang menyimpan kekuatan, dan setiap nasihat sederhana yang justru selalu menguatkan. Untuk pengorbanan yang tidak pernah dihitung, dan kasih sayang yang tidak pernah habis meski tidak selalu terucap. Mama adalah alasan terbesar penulis untuk terus melangkah, menyelesaikan apa yang sudah dimulai, dan tidak menyerah meski jalan terasa berat. Tanpa restu dan cinta Mama, mungkin skripsi ini hanya akan menjadi rencana tanpa akhir. Terima kasih, Mama atas segalanya. Semoga hasil kecil ini bisa menjadi awal dari keberhasilan yang lebih besar, yang suatu hari bisa penulis persembahkan sepenuhnya untuk Mama.
7. Seluruh dosen di Program Studi Teknik Informatika, terima kasih yang sebesar-besarnya penulis ucapkan atas ilmu, bimbingan, dan inspirasi yang telah diberikan selama masa studi. Setiap pertemuan di ruang kelas, setiap penjelasan yang penuh semangat, dan bahkan setiap tugas yang dulu terasa berat, kini terasa sangat berarti. Dari Bapak dan Ibu dosen, penulis belajar bukan hanya tentang teori dan konsep, tapi juga tentang tanggung jawab, integritas, dan semangat untuk terus berkembang. Segala dedikasi dan ketulusan dalam mendidik telah menjadi fondasi penting dalam proses penulis

menyelesaikan skripsi ini. Semoga Allah SWT membalas segala kebaikan dan ilmu yang telah Bapak dan Ibu berikan dengan keberkahan yang tiada putus.

8. Terima kasih yang tulus penulis sampaikan kepada Tim Penguji Skripsi yang telah meluangkan waktu, tenaga, dan perhatian untuk menelaah serta memberikan masukan terhadap skripsi ini. Terima kasih atas pertanyaan, kritik, dan saran yang sangat membangun, yang tidak hanya memperbaiki isi skripsi penulis, tetapi juga memperluas cara pandang penulis dalam melihat permasalahan yang penulis teliti. Kehadiran Bapak dan Ibu sebagai penguji memberikan pengalaman berharga yang akan selalu penulis ingat dalam perjalanan akademik dan kehidupan penulis ke depan.
9. Untuk Gurlly, terima kasih telah menjadi bagian dari perjalanan ini. Bersama kalian, ruang kelas tidak lagi terasa membosankan, dan setiap tugas berat jadi lebih ringan karena tawa, cerita, dan semangat yang kita bagi. Kita mungkin pernah mengeluh bersama, begadang bersama, bahkan nyaris menyerah bersama tapi pada akhirnya, kita tetap berdiri, tetap berjuang, dan terus saling menguatkan. Kalian bukan hanya rekan akademik, tapi juga saudara yang lahir dari perjuangan yang sama. Untuk sahabat-sahabat terbaikku, terima kasih sudah jadi tempat pulang ketika hari terasa berat. Terima kasih atas obrolan larut malam yang sering kali tidak ada hubungannya dengan skripsi, tapi justru menyelamatkan kewarasan. Terima kasih sudah mendengarkan, menyemangati, bahkan menegur saat aku mulai

hilang arah. Kehadiran kalian adalah hadiah yang paling berharga dalam proses ini. Kalau bukan karena kalian, mungkin perjalanan ini akan terasa jauh lebih sunyi.

10. Terima kasih juga penulis sampaikan kepada teman-teman satu bimbingan yang telah menjadi teman seperjalanan dalam menghadapi berbagai proses bimbingan, diskusi, serta berbagi semangat di tengah kesibukan penyusunan skripsi ini. Kebersamaan dalam berbagi pengalaman, saling menguatkan, dan memberikan masukan satu sama lain menjadi bagian berharga yang tidak terlupakan dalam perjalanan ini. Secara khusus, penulis mengucapkan terima kasih kepada teman satu bimbingan, yang selalu hadir memberikan dukungan, semangat, serta menjadi teman bertukar pikiran di saat menghadapi kebingungan dan tantangan. Terima kasih atas kebersamaan, candaan, serta doa yang turut menguatkan penulis hingga akhirnya dapat menyelesaikan skripsi ini.

11. Kepada seseorang yang bersama penulis saat ini yang tidak bisa penulis disebutkan namanya. Terima kasih telah menjadi bagian dari perjalanan hidup penulis. Berkontribusi banyak dalam penulisan karya tulis ini, baik tenaga, waktu, maupun materi kepada penulis. Telah menjadi rumah pendamping dalam segala hal yang menemani, mendukung ataupun menghibur dalam kesedihan, mendengarkan keluh kesah, memberikan semangat untuk pantang menyerah, dan merayakan setiap kemajuan kecil yang mungkin terlihat sepele, tapi

sangat berarti bagi penulis. Semoga Allah selalu memberi keberkahan dalam segala hal yang kita lalui.

12. Dan tak lupa, penulis juga ingin mengucapkan terima kasih kepada diri saya sendiri yang telah bertahan melalui segala tekanan, kelelahan, dan keraguan. Terima kasih telah memilih untuk tidak menyerah, terus belajar, dan percaya bahwa segala usaha ini akan bermakna pada waktunya. Terima kasih telah mempercayai proses, bahkan ketika hasilnya belum terlihat. Terima kasih telah mencintai dirimu sendiri cukup dalam untuk terus berjuang, meski kadang tak ada yang melihat, tak ada yang tahu, dan tak ada yang paham betapa kerasnya semuanya. Kau tidak sempurna, tapi kau luar biasa. Kau tidak selalu kuat, tapi kau tidak pernah benar-benar berhenti. Dan itu cukup untuk disebut luar biasa. Skripsi ini mungkin hanyalah satu bab dalam hidup, tapi perjuangan di baliknya akan selamanya jadi bukti bahwa aku mampu melewati sesuatu yang dulu kupikir tak mungkin. Jadi, untuk diriku sendiri terima kasih, karena sudah sampai di titik ini. Terima kasih, karena tidak menyerah. Terima kasih, karena sudah menjadi rumah yang utuh bagi segala luka, usaha, dan harapan. Ini semua bukan akhir, ini adalah awal dari versi diriku yang jauh lebih tangguh.

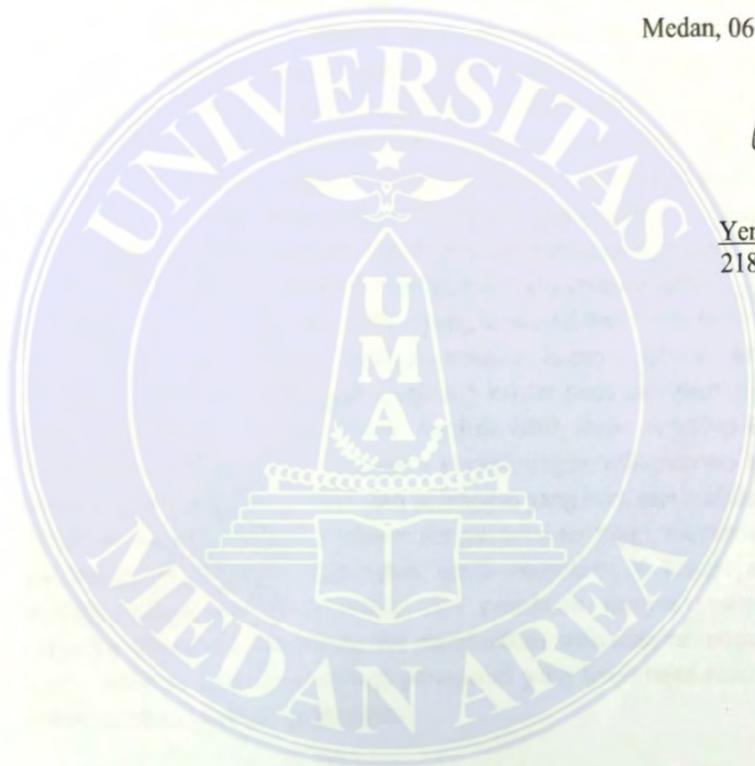
Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, baik dari segi isi, metode, maupun penyajian. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi penyempurnaan karya ini kedepannya.

Akhir kata, semoga skripsi ini dapat memberikan manfaat dan menjadi tambahan referensi dalam penelitian lebih lanjut bagi pihak-pihak yang membutuhkan.

Medan, 06 Maret 2025



YeniSofhiMeliani  
218160008



## ABSTRAK

Pelanggan memiliki peran yang sangat penting dalam keberlangsungan sebuah bisnis. Dimana, dengan adanya pelanggan, bisnis dapat terus berkembang dan beradaptasi dengan kebutuhan pasar yang dinamis. Oleh karena itu, pemahaman tentang tingkat loyalitas pelanggan menjadi hal yang perlu dipertimbangkan bagi perusahaan untuk meningkatkan strategi pemasaran. Penelitian ini bertujuan untuk mengelompokkan pelanggan yang loyal berdasarkan pendekatan RFM (*Recency, Frequency, Monetary*). Metode yang digunakan dalam penelitian ini adalah algoritma *Gaussian Mixture Model* (GMM), yang diimplementasikan dengan variasi konfigurasi *covariance* untuk meningkatkan akurasi pengelompokan. Data yang digunakan dalam penelitian ini merupakan data primer dan sekunder, yang melalui serangkaian tahapan prapemrosesan meliputi pemilihan atribut, pemeriksaan kualitas data, serta normalisasi. Hasil analisis dengan data primer menunjukkan bahwa tipe *covariance* spherical menghasilkan nilai *silhouette score* sebesar 0.8528 dan data sekunder dengan tipe *covariance* spherical menghasilkan nilai *silhouette score* sebesar 0,8101, yang menunjukkan hasil *clustering* yang lebih baik dibandingkan dengan tipe *covariance* lainnya. Model klastering yang dihasilkan membagi pelanggan menjadi 5 klaster pada data primer dan 11 klaster pada data sekunder secara umum. Analisis lebih lanjut terhadap ketiga dimensi RFM menunjukkan bahwa terdapat 5 klaster dengan nilai *recency* terendah, yang berarti klaster-klaster ini terdiri dari pelanggan yang baru saja melakukan transaksi. Untuk *frequency*, terdapat 5 klaster dengan nilai tertinggi, menandakan intensitas pembelian yang tinggi. Sedangkan untuk *monetary*, 4 klaster memiliki nilai tertinggi dalam jumlah transaksi dari penelitian dengan menggunakan data sekunder. Hasil dari penelitian ini memberikan wawasan berharga bagi pemilik bisnis untuk merumuskan strategi pemasaran yang lebih tepat sasaran dan efektif dalam mempertahankan pelanggan.

**Kata Kunci:** Loyalitas, Pelanggan, RFM, Klaster, *Gaussian Mixture Model*.

## ABSTRACT

*Customers play a very important role in the sustainability of a business. With customers, businesses could continue to grow and adapt to dynamic market needs. Therefore, understanding the level of customer loyalty is something that needed to be considered by companies to improve marketing strategies. This research aimed to segment loyal customers based on the RFM (Recency, Frequency, Monetary) approach. The method used in this research was the Gaussian Mixture Model (GMM) algorithm, which was implemented with variations in covariance configurations to improve clustering accuracy. The data used in this research were primary and secondary data, which went through a series of preprocessing stages including attribute selection, data quality checking, and normalization. The analysis results with primary data showed that the spherical covariance type produced a silhouette score of 0.8528, and secondary data with spherical covariance type produced a silhouette score of 0.8101, indicating better clustering results compared to other covariance types. The resulting clustering model divided customers into 5 clusters in primary data and 11 clusters in secondary data in general. Further analysis of the three RFM dimensions showed that there were 5 clusters with the lowest recency values, meaning these clusters consisted of customers who had recently made transactions. For frequency, there were 5 clusters with the highest values, indicating a high purchase intensity. As for monetary, 4 clusters had the highest transaction values from the research using secondary data. The results of this research provided valuable insights for business owners to formulate more targeted and effective marketing strategies in retaining customers.*

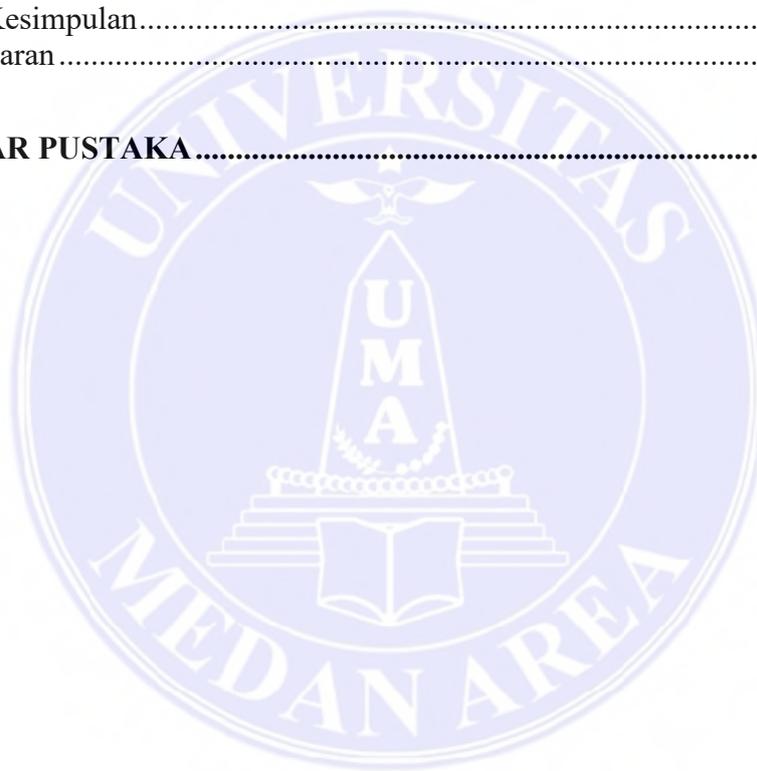
**Keywords:** *Loyalty, Customer, RFM, Cluster, Gaussian Mixture Model.*



## DAFTAR ISI

<b>HALAMAN PENGESAHAN</b> .....	Error! Bookmark not defined.
<b>HALAMAN PERNYATAAN</b> .....	Error! Bookmark not defined.
<b>RIWAYAT HIDUP</b> .....	<b>iv</b>
<b>KATA PENGANTAR</b> .....	<b>v</b>
<b>ABSTRAK</b> .....	<b>xii</b>
<b>ABSTRACT</b> .....	Error! Bookmark not defined.
<b>DAFTAR ISI</b> .....	<b>xiv</b>
<b>DAFTAR TABEL</b> .....	<b>xvi</b>
<b>DAFTAR GAMBAR</b> .....	<b>xvii</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	3
1.3. Batasan Masalah.....	3
1.4. Tujuan Penelitian.....	4
1.5. Manfaat Penelitian.....	4
<b>BAB II TINJAUAN PUSTAKA</b> .....	<b>5</b>
2.1. Loyalitas Pelanggan .....	5
2.2. RFM ( <i>Recency, Frequency, Monetary</i> ).....	6
2.3. Klustering .....	7
2.4. GMM.....	7
2.5. <i>Silhouette Index</i> .....	9
2.6. Penelitian Terdahulu.....	10
<b>BAB III METODOLOGI PENELITIAN</b> .....	<b>12</b>
3.1. Alat dan Bahan .....	12
3.2. Tahapan Penelitian .....	12
3.2.1. Tinjauan Literatur.....	13
3.2.2. Pengumpulan Data .....	14
3.2.3. Prapemrosesan.....	18
3.2.4. EDA ( <i>Exploratory Data Analysis</i> ).....	20
3.2.5. Implementasi .....	21
3.2.6. Evaluasi .....	22
3.3. Perhitungan Manual .....	22
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....	<b>29</b>
4.1. Data yang telah dikumpulkan.....	29
4.2. Prapemrosesan.....	30
4.2.1. Pengecekan Data Null .....	30
4.2.2. Menghitung RFM.....	31
4.2.3. Pemilihan Fitur/Atribut .....	32
4.2.4. Normalisasi .....	32

4.3.	EDA.....	34
4.3.1.	Statistik Deskriptif .....	34
4.3.2.	Distribusi Data pada Atribut .....	35
4.3.3.	Violin .....	39
4.3.4.	Scatter Plot .....	42
4.3.5.	Matriks Korelasi.....	43
4.4.	Implementasi GMM pada Data Primer .....	44
4.4.1.	Menentukan Klaster Pelanggan pada Data Primer.....	45
4.4.2.	Evaluasi pada Data Primer .....	45
4.5.	Implementasi GMM pada Data Sekunder .....	51
4.6.	Pembahasan .....	58
<b>BAB V KESIMPULAN DAN SARAN .....</b>		<b>59</b>
5.1.	Kesimpulan.....	59
5.2.	Saran .....	60
<b>DAFTAR PUSTAKA .....</b>		<b>61</b>



## DAFTAR TABEL

	Halaman
Tabel 2.1. Penelitian Terdahulu .....	10
Tabel 3.1. Rangkuman Data Sekunder Penelitian .....	16
Tabel 3.2. Rangkuman Data Primer Penelitian .....	17
Tabel 3.3. Data Sekunder Penelitian setelah dilakukan <i>Transformasi</i> .....	17
Tabel 3.4. Atribut yang digunakan pada Data Primer .....	18
Tabel 3.5. Atribut yang digunakan pada Data Sekunder.....	18
Tabel 3.6. Data Contoh RFM .....	22
Tabel 3.7. Hasil Data Contoh RFM.....	24
Tabel 3.8. Data Contoh GMM .....	24
Tabel 4.1. Rangkuman Data Primer .....	29
Tabel 4.2. Rangkuman Data Skunder.....	29
Tabel 4.3. Menghitung RFM pada Data Sekunder.....	31
Tabel 4.4. Atribut yang dipilih Pada Data Primer.....	32
Tabel 4.5. Atribut yang dipilih Pada Data Sekunder.....	32
Tabel 4.6. Hasil Normalisasi Data Primer.....	33
Tabel 4.7. Hasil Normalisasi Data Sekunder .....	33
Tabel 4.8. Statistik Deskriptif Data Primer pada RFM .....	34
Tabel 4.9. Statistik Deskriptif Data Sekunder pada RFM.....	34
Tabel 4.10. Distribusi <i>Gender</i> Pada Data Primer.....	35
Tabel 4.11. Distribusi <i>Gender</i> Pada Data Sekunder.....	36
Tabel 4.12. Distribusi Usia Pada Data Primer .....	36
Tabel 4.13. Distribusi Usia Pada Data Sekunder .....	37
Tabel 4.14. Distribusi <i>Product Category</i> pada Data Primer .....	38
Tabel 4.15. Distribusi <i>Product Category</i> pada Data Sekunder .....	39
Tabel 4.16. Matriks Korelasi pada Data Primer.....	43
Tabel 4.17. Matriks Korelasi pada Data Sekunder.....	44
Tabel 4.18. Hasil Klaster Pelanggan Menggunakan Data Primer.....	45
Tabel 4.19. Hasil <i>Silhouette Score</i> pada Data Primer .....	46
Tabel 4.20. Klaster Berdasarkan <i>Recency</i> Menggunakan Data Prime.....	46
Tabel 4.21. Klaster Berdasarkan <i>Frequency</i> pada Data Primer .....	47
Tabel 4.22. Klaster Berdasarkan <i>Monetary</i> pada Data Primer.....	48
Tabel 4.23. Pemberian Label Hasil Klaster pada Data Primer.....	50
Tabel 4.24. Hasil Klaster Pelanggan Menggunakan Data Sekunder .....	51
Tabel 4.25. Hasil <i>Silhouette Score</i> pada Data Sekunder .....	52
Tabel 4.26. Klaster Berdasarkan <i>Recency</i> Menggunakan Data Sekunder .....	53
Tabel 4.27. Klaster Berdasarkan <i>Frequency</i> pada Data Sekunder.....	55
Tabel 4.28. Klaster Berdasarkan <i>Monetary</i> pada Data Sekunder .....	56
Tabel 4.29. Pemberian Label Hasil Klaster pada Data Sekunder .....	57

## DAFTAR GAMBAR

	Halaman
Gambar 3. 1. Flowchart.....	13
Gambar 4.1. Pengecekan Nilai Null.....	31
Gambar 4.2. Bukti Tidda Nilai Null.....	31
Gambar 4.3. Diagram Distribusi Gender pada Data Primer .....	35
Gambar 4.4. Diagram Distribusi Gender pada Data Sekunder .....	36
Gambar 4.5. Diagram Distribusi Usia Pada Data Primer.....	37
Gambar 4.6. Diagram Distribusi Usia Pada Data Sekunder .....	37
Gambar 4.7. Diagram Distribusi <i>Product Category</i> pada Data Primer .....	38
Gambar 4.8. Diagram Distribusi <i>Product Category</i> pada Data Sekunder .....	39
Gambar 4.9. Grafik Violin untuk pada Data Sekunder.....	40
Gambar 4.10. Grafik Violin untuk pada Data Sekunder.....	41
Gambar 4.11. Grafik Violin untuk pada Data Sekunder.....	42
Gambar 4.12. Scatter Plot pada Data Sekunder .....	43
Gambar 4.13. Matriks Korelasi pada Data Primer .....	43
Gambar 4.14. Matriks Korelasi pada Data Sekunder.....	44
Gambar 4.15. Klaster berdasarkan <i>Recency</i> pada Data Primer.....	47
Gambar 4.16. Klaster berdasarkan <i>Frequency</i> pada Data Primer.....	48
Gambar 4.17. Klaster berdasarkan <i>Monetary</i> Pada Data Primer .....	49
Gambar 4.18. Scatter Plot Hasil Klaster .....	52
Gambar 4.19. Klaster berdasarkan <i>Recency</i> pada Data Sekunder.....	54
Gambar 4.20. Klaster berdasarkan <i>Frequency</i> pada Data Sekunder .....	55
Gambar 4.21. Klaster berdasarkan <i>Monetary</i> Pada Data Sekunder .....	56

# BAB I PENDAHULUAN

## 1.1.Latar Belakang

Loyalitas pelanggan adalah seorang pelanggan yang mampu senantiasa membeli barang atau jasa dengan jangka yang cukup panjang (Gultom dkk., 2020). Dalam sebuah bisnis loyalitas pelanggan sangat penting dan dianggap sebagai aset yang berpengaruh besar terhadap suatu usaha atau bisnis (Pratondo dkk., 2024). Pelanggan juga dianggap sebagai nyawa bagi setiap pembisnis atau usaha, untuk mempertahankan seorang pelanggan, pemilik usaha atau pelaku bisnis harus melakukan yang terbaik untuk pelanggan agar tidak kabur (Dahlan dkk., 2022).

Untuk dapat mengetahui dan menentukan loyalitas pelanggan dapat dilakukan dengan menerapkan teknik klasterisasi (Darmawan, 2022). Teknik ini digunakan sebagai identifikasi untuk mengetahui karakteristik pelanggan dan dapat dikelompokkan berdasarkan karakteristik tersebut (Widiyanto & Witanti, 2021). Dengan melakukan teknik ini dapat diketahui pengelompokan loyalitas pelanggan berdasarkan nilai transaksi, frekuensi dan lain-lain (Aini dkk., 2024). Klastering adalah suatu metode pempartisi dengan melakukan pengelompokan sesuai dengan karakteristik tertentu, yang dimana karakteristik tersebut dikelompokkan dengan karakteristik yang memiliki kesamaan dan dikumpulkan ke dalam klaster yang sama (Purba dkk., 2021). klastering adalah suatu metode dalam teknik pengelompokan data atau objek yang dilakukannya dengan mencari informasi kemudian dikelompokkan dengan menjelaskan data objek dengan relasinya (Khaerunnisa, 2022).

Salah satu model yang dapat digunakan adalah GMM (*Gaussian Mixture Model*). GMM adalah pemodelan untuk mengetahui suatu karakteristik sebuah objek dengan menggunakan asumsi agar data tersebut dapat mengikuti karakteristik dengan distribusi gaussian dari parameter yang tidak diketahui (Nurhalizah dkk., 2024). GMM pemodelan yang dilakukan dengan statistika dan dapat mengestimasi sebuah probabilitas dari setiap bobot yang di distribusikan oleh *Gaussian*. GMM sendiri berbeda dengan algoritma pengklasteran lainnya seperti k-means. K-Means adalah teknik pengelompokan yang sederhana dan cepat, namun masih memiliki kelemahan dalam menangkap keberagaman data. Sedangkan, GMM mampu mendeteksi pola yang kompleks dan mengelompokkannya menjadi komponen homogen atau seragam dan memiliki konsistensi dalam kumpulan data (Patel & Kushwaha, 2020). Keunggulan dari GMM adalah pemodelan yang mampu membuat sebuah model lebih dari satu *Gaussian* pada suatu data set. Dimana, *Gaussian* digunakan untuk menggambarkan sejumlah besar titik data mengelompok bersama dan memungkinkan perhitungan yang efisien dalam analisis statistik. (Adhinata dkk., 2021).

Terdapat suatu metode yang dapat digunakan untuk mengklaster pelanggan, yaitu RFM yang terdiri dari R (*Recency*) adalah jarak dan waktu dari transaksi pelanggan dilakukan dengan selisih pada tanggal yang telah ditentukan (Prasetyo dkk., 2020). F (*Frequency*) adalah jumlah dari transaksi yang dilakukan oleh pelanggan selama waktu yang telah ditentukan (Febriani & Putri, 2020). M (*Monetary*) adalah pengukuran jumlah dari besaran transaksi yang telah dilakukan oleh pelanggan dengan waktu yang telah ditetapkan, semakin banyak jumlah yang

dikeluarkan pelanggan makan *monetary* juga akan semakin besar (Firmansah dkk., 2021).

Dari penjelasan yang telah dipaparkan, menentukan loyalitas pelanggan merupakan hal yang cukup penting bagi bisnis dan perusahaan. Namun, penentuan yang dilakukan haruslah menggunakan model yang tepat guna mendapatkan hasil yang maksimal. Sehingga penelitian ini akan melakukan klasterisasi loyalitas pelanggan dengan menggunakan GMM berdasarkan nilai RFM.

### 1.2. Rumusan Masalah

Bagaimana mengklaster pelanggan yang memiliki loyalitas dengan menggunakan GMM berdasarkan RFM.

### 1.3. Batasan Masalah

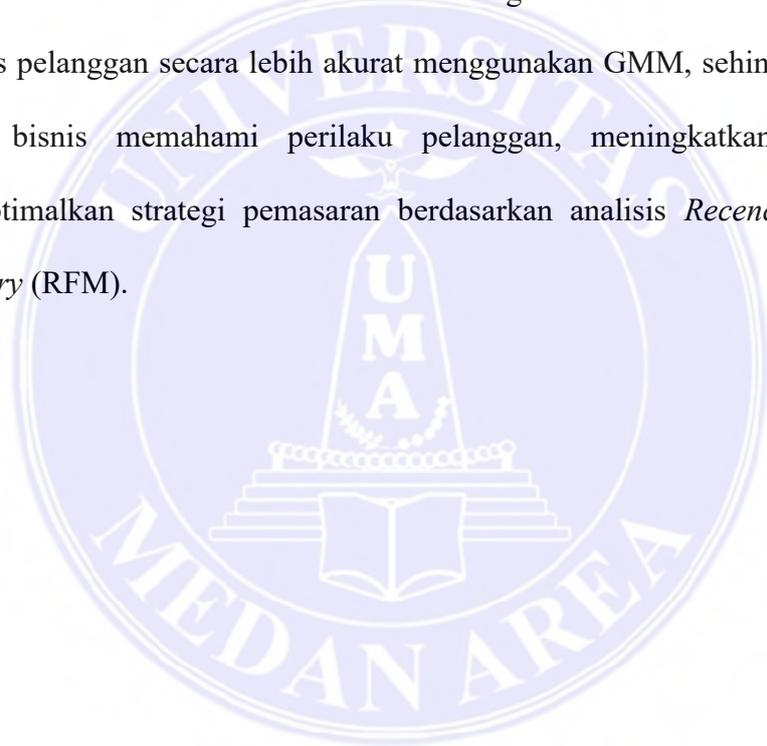
1. Penelitian ini berfokus kepada penggunaan algoritma GMM untuk mengklaster pelanggan berdasarkan tingkat loyalitas.
2. Data yang digunakan adalah data sekunder yang berasal dari Kaggle
3. Atribut yang digunakan adalah RFM.
4. Evaluasi yang digunakan untuk performa model adalah *Silhouette Index*.

#### **1.4. Tujuan Penelitian**

Penelitian ini bertujuan untuk mengklasterisasi agar dapat menentukan loyalitas pelanggan dan pelaku bisnis mengetahui pelanggan mana yang loyal terhadap produk atau jasa, dengan menggunakan GMM.

#### **1.5. Manfaat Penelitian**

Penelitian ini bermanfaat untuk mengidentifikasi dan mengklasterisasi loyalitas pelanggan secara lebih akurat menggunakan GMM, sehingga membantu pelaku bisnis memahami perilaku pelanggan, meningkatkan retensi dan mengoptimalkan strategi pemasaran berdasarkan analisis *Recency*, *Frequency*, *Monetary* (RFM).



## BAB II TINJAUAN PUSTAKA

### 2.1.Loyalitas Pelanggan

Loyalitas adalah komitmen pelanggan kepada pelaku bisnis terhadap barang atau pun jasa. Loyalitas pelanggan juga dapat dilihat dari perilaku dengan menunjukan pembelian barang secara berulang kali dan merekomendasikan barang atau jasa tersebut ke teman atau kerabat. Untuk mengukur indikator terhadap loyalitas pelanggan dapat dilihat dengan pembelian barang secara berulang-ulang dengan teratur (Samara & Susanti, 2023). Dari pandangan psikologis loyalitas pelanggan juga dikatakan sebagai behaviour, Dimana loyalitas berhubungan dengan keyakinan kepada suatu produk atau jasa (Pramudita, 2020).

Loyalitas pelanggan juga memiliki ciri-ciri untuk menentukan keloalitan pelanggan, yaitu seperti Loyalitas Kognitif dimana keyakinan konsumen harus membuat mereka menganggap merek tersebut paling unggul dibanding pesaingnya. Loyalitas didasarkan pada pengetahuan konsumen tentang merek. Loyalitas Afektif yaitu konsumen harus lebih menyukai merek dibanding pesaingnya, sehingga muncul preferensi emosional yang kuat. Sikap ini terbentuk dari ekspektasi sebelum konsumsi dan pengalaman setelahnya. Loyalitas pada tahap ini lebih didasarkan pada perasaan, bukan hanya pikiran, sehingga lebih sulit berubah. Loyalitas Konatif yaitu konsumen harus memiliki niat membeli merek tertentu, bukan yang lain, saat membuat Keputusan. Ini mencerminkan loyalitas yang didasarkan pada komitmen kuat untuk membeli. Dan Loyalitas Tindakan yaitu niat yang didorong oleh motivasi menciptakan kesiapan bertindak dan keinginan

untuk mengatasi hambatan. Tindakan terjadi saat kedua kondisi ini terpenuhi. Loyalitas berkembang dari tahap kognitif, efektif, konatif hingga akhirnya menjadi loyalitas dalam tindakan (Agus dkk., 2024).

## 2.2.RFM (*Recency, Frequency, Monetary*)

Dalam menganalisis perilaku pelanggan dengan melakukan prediksi berdasarkan perilaku yang ada pada data set, dapat menggunakan pemodelan yang berbasis perilaku yaitu model RFM. Model RFM ini adalah metode pengukuran hubungan antara pelaku dan pelanggan (Pailan dkk., 2020).

Pengaplikasian Model RFM ini tertuju kepada database pemasaran dan juga sebagai tools pada teknik untuk membangun pemasaran yang baik. Pemodelan RFM (*Recency, Frequency, Monetary*) memiliki penjelasan masing-masing. *Recency* adalah waktu dari jumlah hari antara akhir periode analisis dengan terjadinya transaksi terakhir pelanggan dilakukan. *Frequency* adalah jumlah dari seberapa sering pelanggan melakukan transaksi dan seberapa kuat hubungan antara pelanggan dengan pelaku. *Monetary* adalah seberapa banyak nilai transaksi yang telah dilakukan oleh pelanggan (Wardani dkk., 2020).

Rumus:

$$R_i = t_p - t_i \quad (2.1)$$

$$F = \text{Jumlah total transaksi dalam periode tertentu} \quad (2.2)$$

$$M = \text{Total nilai transaksi dalam periode tertentu} \quad (2.3)$$

### 2.3. Klustering

Klustering adalah pembelajaran mesin yang digunakan sebagai pengelompokan dengan membuat data menjadi beberapa kelompok atau pun kluster. Pada klustering ini juga memiliki tujuan dalam menemukan struktur yang sebelumnya tidak diketahui (Nahjan dkk., 2023). Dalam metode klustering ini dapat digunakan dengan salah satu metode data mining atau penggalian data. Proses dari klustering ini, yaitu sebagai partisi dari satu set objek ke dalam kelas-kelas atau kelompok yang di sebut dengan kluster (Prastiwi dkk, 2022).

### 2.4. GMM

GMM adalah model yang memiliki komponen fungsi-fungsi *Gaussian* yang terdiri dari *threshold*. Ada pun metode yang efektif seperti *mixture* model pada perubahan model yang bergerak lambat, dengan fleksibilitas yang lebih dan telitih untuk memodelkan statistik dari data atau pun visualisasikan suatu *dynamic scane*. GMM juga dikatakan sebagai model statistik yang distribusikan dengan probalitas dapat dinilai dari bobot pada distribusi *Gaussian*. Dalam perhitungan GMM ini sangat tepat baik dengan parameter atau pun tidak dengan parameter (Sidabutar, 2020).

Untuk membuat distribusi data yang kompleks diperlukan model statistik yaitu GMM. Model ini terdiri dari beberapa kumpulan komponen distribusi *Gaussian* yang berbeda-beda. Dalam bidang klustering, *dimensionality*, pemodelan distribusi data, dan lain-lain dapat menggunakan GMM (Faidah dkk., 2024). GMM mengasumsi bahwa setiap distribusi *Gaussian* mewakili satu kelompok. Setiap *Gaussian* ditentukan oleh kombinasi nilai rata-rata dan varians. Tujuan klastring

dengan GMM adalah menemukan parameter model yang paling cocok dengan data. Ada beberapa model yang digunakan untuk klastring dengan memperhatikan bentuk geometris yang dihasilkan oleh komponen *Gaussian*, Dimana setiap komponen memiliki parameter yang berbeda. Fungsi kepadatan peluang untuk distribusi *Gaussian* satu dimensi dapat diwakili dengan persamaan tertentu.

Rumus:

$$f(X|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2.4)$$

Dengan  $\mu$  sebagai rata-rata dan  $\sigma$  sebagai standar deviasi dari suatu distribusi, fungsi kepadatan peluang untuk data multivariat dapat dinyatakan dengan persamaan tertentu yang berlaku untuk lebih dari satu dimensi.

Rumus :

$$f(X|\mu, \Sigma) = \frac{1}{\sqrt{2\pi|\Sigma|}} e^{\left(-\frac{1}{2}(X-\mu)^T \Sigma^{-1}(X-\mu)\right)} \quad (2.5)$$

Dengan  $\mu$  mewakili rata-rata dari distribusi yang ditampilkan dalam suatu array berdimensi  $d$ ,  $\Sigma$  adalah matriks kovarians dari  $X$ .  $T$  menunjukkan vektor transpose, dan  $-1$  menacu pada invers dan matriks.

Pemilihan model terbaik dalam metode GMM didasarkan pada pendekatan umum menggunakan *Bayes Information Criterion* (BIC). Model dan jumlah kluster terbaik ditentukan berdasarkan nilai BIC tertinggi.

Rumus:

$$BIC_k = 2 \log P(y|\hat{\theta}_k, M_k) - V_k \log(n) \approx 2 \log P(y|M_k) \quad (2.6)$$

Pengelompokan adalah teknik yang digunakan untuk membagi data yang tidak memiliki label ke dalam kelompok-kelompok berdasarkan kemiripannya, tanpa banyak bimbingan. Objek-objek dalam satu kelompok memiliki karakteristik yang mirip satu sama lain, sementara objek dalam kelompok yang berbeda memiliki karakteristik yang berbeda. Pengelompokan juga merupakan bagian dari pembelajaran mesin, terutama dalam kategori pembelajaran tanpa pengawasan (Oyewole & Thopil, 2023).

### 2.5. Silhouette Index

*Silhouette index* adalah metode yang kuat untuk menilai hasil pengelompokan, selain metode *elbow*. Metode ini tidak memerlukan set pelatihan, sehingga lebih cocok untuk tugas pengelompokan. Pengelompokan yang ideal, diukur dengan *silhouette index*, didasarkan pada perbedaan antara rata-rata jarak antar objek dalam satu klaster dan jarak minimum antar klaster. Index ini bisa membantu mengidentifikasi apakah objek dalam data klaster sudah berada di kelompok yang tepat atau berada di area transisi antara klaster yang berbeda (Lubis dkk., 2023).

Rumus:

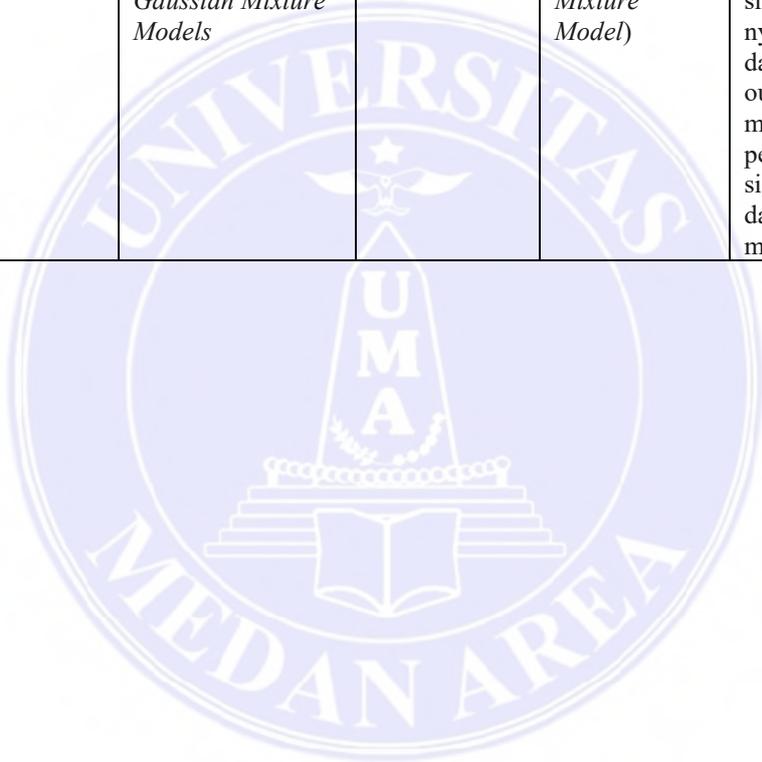
$$SI(i) = \frac{b(i) - a(i)}{\max\{a(i) - b(i)\}} \quad (2.7)$$

## 2.6. Penelitian Terdahulu

**Tabel 2.1.** Penelitian Terdahulu

Penulis (tahun)	Judul	Objek	Metode	Hasil
(Patel & Kushwaha, 2020)	<i>Clustering Cloud Workloads: K-Means vs Gaussian Mixture Model</i>	Beban Kerja Cloud	K-Means dan Gaussian Mixture Model (GMM)	Pada metode K-Means memberikan kluster yang lebih sederhana dan juga abstrak terkait dengan penggunaan sumber daya, sedangkan GMM mampu menghasilkan kluster dengan batas-batas pengguna yang lebih rinci, dengan mencerminkan pola heterogenitas yang lebih kompleks.
(Sekaran dkk., 2020)	<i>Deep learning convolutional neural network (CNN) With Gaussian mixture model for predicting pancreatic cancer</i>	Kanker Pankreas	CNN dan GMM	Model ini berhasil memprediksi penyebaran kanker pankreas dengan Tingkat akurasi yang lebih tinggi dibandingkan metode konvensional. Hasil eksperimen menunjukkan bahwa setelah diagnosis dan perawatan ukuran tumor menurun secara signifikan, meningkatkan peluang pemulihan pasien.
(Mahmood dkk., 2023)	<i>Optimizing Clustering Algorithms for Anti-Microbial Evaluation Data: A Majority Score-Based Evaluation of K-Means, Gaussian Mixture Model, and Multivariate T-</i>	Anti-Mikroba	K-Means, GMM, dan Multivariate T-Distribution Mixtures	K-Means dan Multivariate T-Distribution Mixtures menghasilkan tiga kluster optimal untuk aktivitas antibakteri dan lima kluster untuk data MIC bakteri. GMM

	<i>Distribution Mixtures</i>			memberikan hasil terbaik untuk nilai minimum CVIs, sedangkan <i>K-Means</i> dan <i>Multivariate T-Distribution</i> lebih unggul dalam nilai Silhouette Width dan Dunn Index
(TU dkk., 2024)	<i>Change Point Detection Via Gaussian Mixture Models</i>	Titik perubahan	GMM ( <i>Gaussian Mixture Model</i> )	Pada algoritma ini berhasil mendeteksi titik perubahan secara efektif pada data simulasi dan nyata, unggul dalam mengatasi outlier dan mengidentifikasi perubahan signifikan pada data univariat dan multivariat.



## **BAB III**

### **METODOLOGI PENELITIAN**

#### **3.1. Alat dan Bahan**

Dalam menyusun tugas akhir ini, diperlukan perangkat pendukung yang dapat membantu, baik dari segi perangkat kerja (*Hardware*) dan perangkat lunak (*Software*). Berikut adalah perangkat keras dan lunak yang akan digunakan :

1. Perangkat Keras:

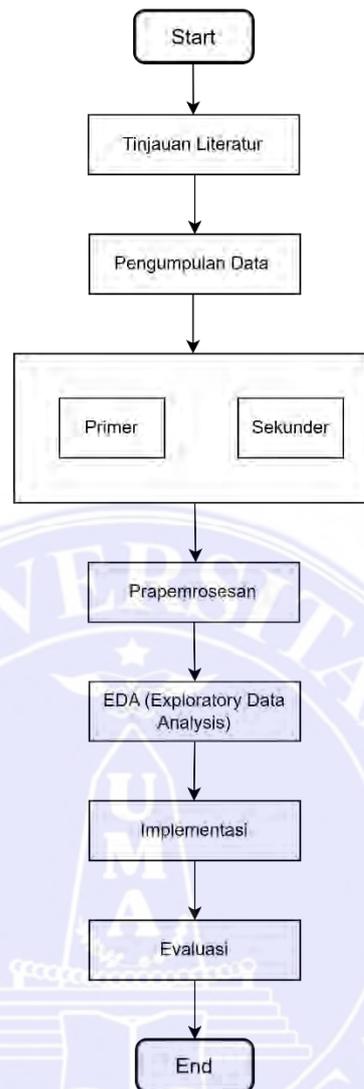
- a. *Laptop ThinkPad Windowa 10*
- b. *Processor Intel Core i3 CPU @ 2.40GHz*
- c. *Memory (RAM) 12 GB*

2. Perangkat Lunak:

- a. *Microsoft word*
- b. *Microsoft excel*
- c. *Google Colab*

#### **3.2. Tahapan Penelitian**

Pada penelitian ini memiliki beberapa tahapan yang harus dilakukan dan dimulai. Penelitian ini dimulai dari peninjauan literatur sampai dengan tahapan evaluasi hasil dari penelitian ini. Tahapan penelitian tersebut dapat dilihat dari Gambar 3.1 di bawah ini.



Gambar 3. 1. Flowchart

### 3.2.1. Tinjauan Literatur

Tinjauan literatur pada penelitian ini dilakukan untuk mencari referensi dan rujukan yang berkaitan dengan penelitian ini. Dengan adanya referensi penelitian ini dapat dikembangkan. Sehingga dalam penelitian ini dapat memberikan kontribusi yang baik dalam keberlanjutan untuk objek pada pembahasan penelitian ini. Adapun sumber-sumber yang diambil untuk dijadikan literatur pada penelitian ini

adalah yang berkaitan dengan loyalitas pelanggan, teknik klastrer, RFM, dan GMM dari penelitian terdahulu.

### 3.2.2. Pengumpulan Data

Penelitian ini menggunakan data primer dan sekunder sebagai sumber data. Data primer dikumpulkan dengan cara melakukan survei kepada responden dengan jumlah 47 pelanggan. Data primer terdiri atas berbagai atribut, yakni sejumlah 8 atribut yang diantaranya, “Nama” yang berfungsi untuk mengetahui nama pelanggan yang sudah melakukan transaksi, “Usia” digunakan untuk melihat usia dari pembeli agar dapat menganalisis kelompok usia, sedangkan “Jenis Kelamin” untuk mengetahui pelanggan yang membeli laki-laki atau perempuan, “Nama/Merek produk” digunakan untuk mengetahui produk atau merek apa yang dipakai oleh pelanggan, dan “Seberapa sering” untuk mengukur seberapa sering pelanggan memakai produk tersebut.

Pada data primer, terdapat atribut “Kapan terakhir kali” yang digunakan untuk menggambarkan *Recency*. Kemudian, “Berapa kali anda membeli produk” adalah atribut yang digunakan untuk menggambarkan nilai dari *Frequency*. Sedangkan atribut “Berapa rata-rata jumlah uang” digunakan untuk menggambarkan nilai *Monetary*. Hasil pengumpulan data primer dirangkum dalam Tabel 3.1.

Data sekunder adalah data yang sudah tersedia dan dikumpulkan oleh pihak lain yang kemudian digunakan untuk keperluan penelitian (Trilaksono & Handayani, 2020). Pada penelitian ini, peneliti mengambil dataset, yang dimana data tersebut diambil peneliti melalui *website Kaggle*, dengan tautan

[www.kaggle.com/datasets/mohammadtalib786/retail-sales-dataset](https://www.kaggle.com/datasets/mohammadtalib786/retail-sales-dataset). Jumlah baris data yang digunakan dalam penelitian ini sebanyak 1000 data yang terdiri dari 9 atribut (kolom). Atribut-atribut tersebut terdiri dari “*Transaction id*”, yang merupakan penomoran khusus yang dibuat setiap kali ada transaksi. “*Date*” adalah waktu yang berisi tanggal, bulan dan tahun yang digunakan untuk mencatat kapan data diproses. “*Customer id*” adalah nomor unik yang diberikan ke setiap pelanggan. Atribut “*Gender*” adalah jenis kelamin pelanggan, lalu atribut “*Age*” adalah usia pelanggan. Kemudian, terdapat pula atribut “*Product Category*” yang merupakan pengelompokan jenis produk yang dibeli oleh pelanggan.

Selain itu, atribut “*Quantity*” digunakan untuk menunjukkan jumlah produk yang dibeli oleh pelanggan. Lalu, atribut “*Price per unit*” digunakan untuk mengetahui harga untuk suatu produk. Total amount adalah jumlah akhir yang mencakup semua harga yang dibayar oleh pelanggan. Tabel 3.3 di bawah ini menampilkan rangkuman data dari sumber yang akan diproses pada penelitian ini.

**Tabel 3.1.** Rangkuman Data Primer Penelitian

No	Nama	Usia	Jenis Kelamin	Nama/Merek produk	Seberapa sering	Kapan terakhir kali	Berapa kali anda membeli produk	Berapa rata-rata jumlah uang
1	Cut fara farisya		Perempuan	Skin1004	Setiap 2-3 bulan	kurang dari 1 bulan yang lalu	1 - 2 kali	Rp 100.000 - Rp 300.000
2	Rindi	18-25	Perempuan	Skin1004	Jarang (lebih dari 3 bulan)	2 - 3 minggu yang lalu	1 - 2 kali	Rp 300.000 - Rp 500.000
3	Syahliuca Endrian	18-25	Laki-laki	Kahf	Setiap bulan	2 - 3 minggu yang lalu	Tidak pernah lagi membeli	Rp 100.000 - Rp 300.000
4	Luthfi Chaidar Ikhrum	<18	Laki-laki	Kahf	Jarang (lebih dari 3 bulan)	Lebih dari 3 bulan yang lalu	3 - 5 kali	Kurang dari Rp 100.000
5	Bambang Saktiawan	18-25	Laki-laki	RODEOS	Setiap bulan	2 - 3 minggu yang lalu	1 - 2 kali	Rp 100.000 - Rp 300.000
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
44	Nawan	18-25	Laki-laki	Bio aqua	Setiap bulan	2 - 3 minggu yang lalu	3 - 5 kali	Rp 100.000 - Rp 300.000
45	Andika Silalahi	18-25	Laki-laki	Nivea Men	Setiap bulan	Kurang dari seminggu	3 - 5 kali	Rp 100.000 - Rp 300.000
46	Andre Laksmna Saragih	18-25	Laki-laki	Sabun telepon	Setiap minggu	kurang dari 1 bulan yang lalu	1 - 2 kali	Kurang dari Rp 100.000
47	putri maya gemilang	18-25	Perempuan	Wardah	Setiap bulan	kurang dari 1 bulan yang lalu	1 - 2 kali	Rp 300.000 - Rp 500.000

**Tabel 3.2.** Rangkuman Data Primer Penelitian

<i>Transaction ID</i>	<i>Date</i>	<i>Customer ID</i>	<i>Gender</i>	<i>Age</i>	<i>Product Category</i>	<i>Quantity</i>	<i>Price per Unit</i>	<i>Total Amount</i>
1	24/11/2023	CUST001	Male	34	Beauty	3	50	150
2	27/02/2023	CUST002	Female	26	Clothing	2	500	1000
3	13/01/2023	CUST003	Male	50	Electronics	1	30	30
...	...	...	...	...	...	...	...	...
1000	12/04/2023	CUST1000	Male	47	Electronics	4	30	120

**Tabel 3.3.** Data Sekunder Penelitian setelah dilakukan *Transformasi*

<i>Transaction ID</i>	<i>Date</i>	<i>Customer ID</i>	<i>Gender</i>	<i>Age</i>	<i>Product Category</i>	<i>Quantity</i>	<i>Price per Unit</i>	<i>Total Amount</i>
1	24/11/2023	CUST001	2	34	19	3	50	150
2	27/02/2023	CUST002	1	26	5	2	500	1000
3	13/01/2023	CUST003	2	50	3	1	30	30
...	...	...	...	...	...	...	...	...
1000	12/04/2023	CUST1000	2	47	3	4	30	120

### 3.2.3. Prapemrosesan

Prapemrosesan data adalah tahap awal yang penting dalam analisis, dimana data mentah dibersihkan dan diubah agar siap diproses lebih lanjut. Proses ini melibatkan beberapa proses, yaitu pemilihan fitur atau atribut yang digunakan, *transformasi* data, pengecekan nilai kosong (null) dan normalisasi. Proses pemilihan fitur dilakukan dengan menentukan atribut mana yang relevan untuk digunakan. Tabel 3.4 di bawah ini menampilkan atribut yang dipilih pada penelitian ini.

**Tabel 3. 4.** Atribut yang digunakan pada Data Primer

Atribut	Penggunaan
Nama	<b>X</b>
Usia	<b>X</b>
Jenis Kelamin	<b>X</b>
Nama/Merek_produk	<b>X</b>
R	√
F	√
M	√

Pada data sekunder ini dapat dilihat dari Tabel 3.5 di bawah ini yang menampilkan atribut mana saja yang digunakan.

**Tabel 3. 5.** Atribut yang digunakan pada Data Sekunder

Atribut	Penggunaan
<i>Transaction ID</i>	<b>X</b>
<i>Date</i>	√
<i>Customer ID</i>	√
<i>Gender</i>	<b>X</b>
<i>Age</i>	<b>X</b>
<i>Product Category</i>	<b>X</b>
<i>Quantity</i>	√
<i>Price per Unit</i>	√
<i>Total Amount</i>	√

Atribut yang digunakan pada Tabel 3.5 di atas lebih berfokus pada RFM dan untuk data yang tidak digunakan, seperti *Transaction ID* digunakan pada saat memerlukan pelacakan atau memverifikasi transaksi tertentu. *Customer ID* digunakan untuk menghubungkan hasil analisis dengan pelanggan tertentu. Setelah kluster selesai, kemudian dapat mengidentifikasi pelanggan mana yang masuk ke dalam kluster loyalitas tertinggi. *Gender* digunakan setelah proses kluster RFM selesai. Data gender dapat digunakan untuk memahami karakteristik demografis dari masing-masing kluster pelanggan. *Age* dapat digunakan setelah proses klustering RFM selesai untuk menganalisis demografi setiap kluster, agar dapat dilihat pelanggan paling loyal cenderung berasal dari rentan usia tertentu. *Product Category* tidak langsung dihitung dalam analisis RFM, atribut ini sangat bermanfaat untuk analisis lebih lanjut ini dapat membantu memahami preferensi produk sesuai dengan kluster pelanggan yang telah terbentuk. Untuk atribut yang digunakan, seperti *Date* untuk menghitung *recency*, yang menunjukkan kapan pelanggan terakhir bertransaksi. *Quantity* berperan dalam perhitungan *frequency* dan *monetary*. Jumlah produk yang dibeli mencerminkan seberapa sering pelanggan bertransaksi dan seberapa banyak uang yang telah dikeluarkan. *Price* per unit untuk menghitung total pembelian pelanggan (*monetary*). Dengan mengetahui harga perunit. Total amount digunakan langsung untuk menghitung *monetary*, yaitu total uang yang dikeluarkan pelanggan dalam periode tertentu.

Kemudian, dilakukan proses pengecekan dan penanganan nilai-nilai yang tidak valid seperti null dan "NaN", yang dapat muncul akibat kesalahan saat pengumpulan data. Setelah itu, teknik transformasi seperti *one hot encoding* untuk mengonversi variabel kategori menjadi bentuk numerikal, digunakan agar data

sesuai dengan kebutuhan analisis dan model prediksi seperti pada Tabel 3.3. Transformasi data ini memastikan bahwa hasil analisis menjadi lebih akurat dan dapat diandalkan. Lebih lanjut, proses normalisasi dilakukan untuk merubah skala data, standarisasi untuk membuat data memiliki distribusi yang lebih konsisten. Rumus untuk melakukan proses normalisasi adalah sebagai berikut.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3.1)$$

#### 3.2.4. EDA (*Exploratory Data Analysis*)

EDA adalah proses untuk memahami data dengan lebih baik sebelum analisis lebih lanjut, dengan mencari pola, melihat bagaimana data tersebar, menemukan nilai-nilai yang tidak biasa, dan mendapatkan informasi awal yang membantu dalam memilih metode analisis yang tepat. EDA biasanya melibatkan korelasi, visualisasi data, dan statistik deskriptif untuk membantu dalam memahami data dengan lebih baik.

Korelasi dalam EDA adalah metode yang digunakan untuk mengevaluasi kekuatan hubungan antara dua variable diikuti oleh perubahan pada variable lain. Jika kedua variable bergerak searah, baik naik maupun turun bersama ini disebut korelasi positif. Jika satu variable naik sementara yang lain turun, disebut korelasi negatif. Angka korelasi berkisar antara -1 (hubungan negatif sempurna) hingga 1 (hubungan positif sempurna), dengan 0 berarti tidak ada hubungan sama sekali. Korelasi yang dihasilkan adalah total amount memiliki hubungan yang signifikan

pada *price* per unit, dengan memperoleh hasil 0.85, dengan menggunakan program *python*. Dapat dilihat pada gambar dibawah ini.

Statisti deskriptif dalam EDA adalah cara sederhana untuk menggambarkan data seperti menghitung rata-rata (nilai tengah semua data), median (nilai tengah setelah data diurutkan), modus (nilai yang paling sering muncul), range (selisih nilai terbesar dan terkecil), Standar deviasi (seberapa jauh data tersebar dari rata-rata), seta persentil dan kuartil (membagi data untuk melihat distribusi). Pada table dibawah ini, berikut adalah hasil dari statistik deskriptif.

Visualisasi data dalam EDA adalah cara menampilkan data dalam bentuk grafik, seperti histogram, *scatter plot*, *box plot*, *bar chart*, dan heatmap untuk mempermudah dalam melihat pola, tren, serta hubungan antara variable secara lebih jelas.

### 3.2.5. Implementasi

Implementasi analisis RFM dengan menggunakan GMM dan membandingkanya K-means klastering digunakan untuk mengelompokan pelanggan dengan cara menilai kapan terakhir mereka bertransaksi, seberapa sering mereka berbelanja, dan berapa banyak yang mereka belanjakan, kemudian mengurangi jumlah variable data untuk visualisasi yang lebih sederhana, dan akhirnya membagi pelanggan menjadi klaster-klaster yang berbeda menggunakan RFM, sehingga dapat membantu menyusun pemasaran yang lebih efektif.

### 3.2.6. Evaluasi

*Silhouette index* digunakan mengevaluasi kualitas klaster dengan membandingkan seberapa dekat objek dalam klaster yang sama dan seberapa jauh objek tersebut dari klaster lain yang membantu menentukan apakah data sudah dikelompokkan dengan baik dan menemukan jumlah klaster yang paling sesuai untuk segmentasi.

### 3.3. Perhitungan Manual

Perhitungan manual dilakukan dalam penelitian ini untuk membantu peneliti memahami data dengan lebih baik, memastikan hasil yang diperoleh akurat, serta memberikan kejelasan dan transparansi dalam proses analisis yang dilakukan.

#### 1. Perhitungan pada RFM

**Tabel 3.6.** Data Contoh RFM

Customer ID	Tanggal Terakhir transaksi	Jumlah Transaksi	Total pembelian (RP)
P01	01-01-2024	5	2.000.000
P02	15-12-2023	3	1.000.000
P03	25-11-2023	8	3.500.000
P04	10-01-2024	2	750.000
P05	05-12-2023	6	2.200.000

#### a. Recency

*Recency* mengukur berapa hari jarak transaksi dilakukan oleh pelanggan. Maka semakin kecil jumlah hari, semakin baru juga pelanggan melakukan transaksi.

Rumus:

$$R = \text{Tanggal sekarang} - \text{Tanggal Transaksi Terakhir}$$

Menghitung *recency* sampai dengan tanggal 31 Januari 2024

P01: (31 – 01 – 2024) sampai (01 – 01 – 2024) = 30 hari

P02: (31 – 01 – 2024) sampai (15 – 12 – 2023) = 47 hari

P03: (31 – 01 – 2024) sampai (25 – 11 – 2023) = 67 hari

P04: (31 – 01 – 2024) sampai (10 – 01 – 2024) = 21 hari

P05: (31 – 01 – 2024) sampai (05 – 12 – 2023) = 57 hari

b. *Frequency*

Jika nilai dari *frequency* semakin besar, maka pelanggan tersebut sering bertransaksi

- P01: 5 transaksi
- P02: 3 transaksi
- P03: 8 transaksi
- P04: 2 transaksi
- P05: 6 transaksi

c. *Monetary*

Jika pada *monetary* terdapat nilai yang besar, maka besar pula pengeluaran yang harus dibayar oleh pelanggan.

- P01: 2.000.000
- P02: 1.000.000
- P03: 3.500.000
- P04: 750.000
- P05: 2.200.000

Hasil akhir skor RFM adalah sebagai berikut:

**Tabel 3.7.** Hasil Data Contoh RFM

<i>Customer ID</i>	<i>Recency (Hari)</i>	<i>Frequency</i>	<i>Monetary (RP)</i>
P01	30	5	2.000.000
P02	47	3	1.000.000
P03	67	8	3.500.000
P04	21	2	750.000
P05	57	6	2.200.000

Dapat dilihat dari Tabel 3.5 di atas bahwasannya “pelanggan loyal aktif” terdapat pada (P01, P04), sedangkan “pelanggan loyal tapi tidak aktif” terdapat pada (P03, P05), dan “pelanggan pasif atau tidak loyal” terdapat pada (P02).

## 2. Perhitungan pada GMM

Tabel data yang digunakan untuk perhitungan

**Tabel 3.8.** Data Contoh GMM

<i>Transaction ID</i>	<i>Date</i>	<i>Customer ID</i>	<i>Gender</i>	<i>Age</i>	<i>Product Category</i>	<i>Quantity</i>	<i>Price per Unit</i>	<i>Total Amount</i>
1	24/11/2023	CUST001	2	34	19	3	50	150
2	27/02/2023	CUST002	1	26	5	2	500	1000

Data yang digunakan :

- a.  $Age = 34$ ,  $Total\ Amount = 150$
- b.  $Age = 26$ ,  $Total\ Amount = 1000$

### 1) Inisialisasi awal parameter *Gaussian*

- **Komponen Gaussian 1:**
  - $Mean (\mu_1) = 30$
  - $Variance (\sigma_1^2) = 25 (\sigma_1 = 5)$
  - $Weight (\pi_1) = 0.5$
- **Komponen Gaussian 2:**
  - $Mean (\mu_2) = 40$
  - $Variance (\sigma_2^2) = 16 (\sigma_2 = 4)$

$$\blacksquare \text{ Weight } (\pi_2) = 0.5$$

2) *Expectation step (E-step)*

Menghitung probabilitas keanggotaan setiap data untuk masing-masing komponen *Gaussian*

- Menghitung probabilitas untuk data pertama (*Age* = 34)

*Gaussian 1:*

$$\begin{aligned} P(x|\mu_1, \sigma_1^2) &= \frac{1}{\sqrt{2\pi \cdot 25}} \exp\left(-\frac{(34-30)^2}{2 \cdot 25}\right) \\ &= \frac{1}{\sqrt{50\pi}} \exp\left(-\frac{16}{50}\right) \\ &= \frac{1}{12.53} \times \exp(-0.32) = 0.0797 \end{aligned}$$

*Gaussian 2:*

$$\begin{aligned} P(x|\mu_2, \sigma_2^2) &= \frac{1}{\sqrt{2\pi \cdot 16}} \exp\left(-\frac{(34-40)^2}{2 \cdot 16}\right) \\ &= \frac{1}{\sqrt{32\pi}} \exp\left(-\frac{36}{32}\right) \\ &= \frac{1}{10.03} \times \exp(-1.125) = 0.0324 \end{aligned}$$

- Menghitung probabilitas untuk data kedua (*Age* = 26)

*Gaussian 1:*

$$\begin{aligned} P(x|\mu_1, \sigma_1^2) &= \frac{1}{\sqrt{2\pi \cdot 25}} \exp\left(-\frac{(26-30)^2}{2 \cdot 25}\right) \\ &= \frac{1}{\sqrt{50\pi}} \times \exp\left(-\frac{16}{50}\right) \\ &= \frac{1}{12.53} \times \exp(-0.32) = 0.0797 \end{aligned}$$

*Gaussian 2:*

$$\begin{aligned} P(x|\mu_2, \sigma_2^2) &= \frac{1}{\sqrt{2\pi \cdot 16}} \exp\left(-\frac{(26-40)^2}{2 \cdot 16}\right) \\ &= \frac{1}{\sqrt{32\pi}} \exp\left(-\frac{196}{32}\right) \\ &= \frac{1}{10.03} \times \exp(-6.125) = 0.0002 \end{aligned}$$

- Menghitung total probabilitas

- Data pertama (*Age* = 34):

$$P(x_1) = \pi_1 P(x_1 | \mu_1, \sigma_1^2) + \pi_2 P(x_1 | \mu_2, \sigma_2^2)$$

$$= 0.5 \times 0.0797 + 0.5 \times 0.0324 = 0.03985 + 0.0162 = 0.05605$$

- Data kedua (*Age* = 26):

$$P(x_2) = \pi_1 P(x_2 | \mu_1, \sigma_1^2) + \pi_2 P(x_2 | \mu_2, \sigma_2^2)$$

$$= 0.5 \times 0.0797 + 0.5 \times 0.0002 = 0.03985 + 0.0001 = 0.03995$$

- Probabilitas posterior

- Data pertama (*Age* = 34)

*Gaussian* 1:

$$P(\text{Gaussian 1} | x_1) = \frac{\pi_1 P(x_1 | \mu_1, \sigma_1^2)}{P(x_2)} = \frac{0.5 \times 0.0797}{0.05605} = \frac{0.03985}{0.05605} = 0.711$$

*Gaussian* 2:

$$P(\text{Gaussian 2} | x_1) = \frac{\pi_2 P(x_1 | \mu_2, \sigma_2^2)}{P(x_1)} = \frac{0.5 \times 0.0324}{0.05605} = \frac{0.0162}{0.05605} = 0.289$$

- Data kedua (*Age* = 26)

*Gaussian* 1 :

$$P(\text{Gaussian 1} | x_2) = \frac{\pi_1 P(x_2 | \mu_1, \sigma_1^2)}{P(x_2)} = \frac{0.5 \times 0.0797}{0.03995} = \frac{0.03985}{0.03995} = 0.997$$

*Gaussian* 2:

$$P(\text{Gaussian 2} | x_1) = \frac{\pi_2 P(x_2 | \mu_2, \sigma_2^2)}{P(x_2)} = \frac{0.5 \times 0.0002}{0.03995} = \frac{0.0001}{0.03995} = 0.003$$

### 3) Maximization step

Memperbarui parameter *Gaussian* berdasarkan probabilitas posterior

- Perbarui *Mean* ( $\mu$ )

*Gaussian* 1:

$$\mu'_1 = \frac{P(\text{Gaussian 1} | x_1) \times x_1 + P(\text{Gaussian 1} | x_2) \times x_2}{P(\text{Gaussian 1} | x_1) + P(\text{Gaussian 1} | x_2)}$$

$$\begin{aligned} \mu'_1 &= \frac{0.711 \times 34 + 0.997 \times 26}{0.711 + 0.997} = \frac{24.174 + 25.922}{1.708} \\ &= \frac{50.096}{1.708} = 29.33 \end{aligned}$$

*Gaussian* 2:

$$\mu'_2 = \frac{P(\text{Gaussian } 2|x_1) \times x_1 + P(\text{Gaussian } 2|x_2) \times x_2}{P(\text{Gaussian } 2|x_1) + P(\text{Gaussian } 2|x_2)}$$

$$\mu'_2 = \frac{0.289 \times 34 + 0.003 \times 26}{0.289 + 0.003} = \frac{9.826 + 0.078}{0.292}$$

$$= \frac{9.904}{0.292} = 33.91$$

- Perbarui *Variance* ( $\sigma^2$ )

*Gaussian 1:*

$$\sigma_1'^2 = \frac{P(\text{Gaussian } 1|x_1) \times (x_1 - \mu'_1)^2 + P(\text{Gaussian } 1|x_2) \times (x_2 - \mu'_1)^2}{P(\text{Gaussian } 1|x_1) + P(\text{Gaussian } 1|x_2)}$$

Data:

$$x_1 = 34, x_2 = 26$$

$$\mu'_1 = 29.33$$

$$\sigma_1'^2 = \frac{0.711 \times (34 - 29.33)^2 + 0.997 \times (26 - 29.33)^2}{0.711 + 0.997}$$

$$= \frac{0.711 \times (4.67)^2 + 0.997 \times (-3.33)^2}{1.708}$$

$$= \frac{0.711 \times 21.81 + 0.997 \times 11.09}{1.708}$$

$$= \frac{15.51 + 11.06}{1.708} = \frac{26.57}{1.708} = 15.56$$

*Gaussian 2:*

$$\sigma_2'^2 = \frac{P(\text{Gaussian } 2|x_1) \times (x_1 - \mu'_2)^2 + P(\text{Gaussian } 2|x_2) \times (x_2 - \mu'_2)^2}{P(\text{Gaussian } 2|x_1) + P(\text{Gaussian } 2|x_2)}$$

Data:

$$x_1 = 34, x_2 = 26$$

$$\mu'_2 = 33.91$$

$$\sigma_2'^2 = \frac{0.289 \times (34 - 33.91)^2 + 0.003 \times (26 - 33.91)^2}{0.289 + 0.003}$$

$$\begin{aligned}
 &= \frac{0.289 \times (0.09)^2 + 0.003 \times (-7.91)^2}{0.292} \\
 &= \frac{0.289 \times 0.0081 + 0.003 \times 62.56}{0.292} \\
 &= \frac{0.00234 + 0.18768}{0.292} = \frac{0.19002}{0.292} = 0.651
 \end{aligned}$$

- Menghitung bobot untuk iterasi selanjutnya:

Bobot pada setiap *Gaussian* dihitung berdasarkan rata-rata dari probabilitas posterior untuk setiap data.

$$\begin{aligned}
 \pi'_1 &= \frac{P(\text{Gaussian 1}|x_1) + P(\text{Gaussian 1}|x_2)}{N} \\
 \pi'_1 &= \frac{0.711 + 0.997}{2} = \frac{1.708}{2} = 0.854 \\
 \pi'_2 &= \frac{P(\text{Gaussian 2}|x_1) + P(\text{Gaussian 2}|x_2)}{N} \\
 \pi'_2 &= \frac{0.289 + 0.003}{2} = \frac{0.292}{2} = 0.146
 \end{aligned}$$

- Parameter yang telah diperbarui dengan iterasi
  - *Gaussian 1*:  
 Mean ( $\mu'_1$ ) = 29.33  
 Variance ( $\sigma_1'^2$ ) = 15.56  
 Weight ( $\pi'_1$ ) = 0.854
  - *Gaussian 2*:  
 Mean ( $\mu'_1$ ) = 33.91  
 Variance ( $\sigma_1'^2$ ) = 0.651  
 Weight ( $\pi'_1$ ) = 0.146

## BAB V KESIMPULAN DAN SARAN

### 5.1. Kesimpulan

Penelitian ini bertujuan untuk mengelompokkan pelanggan berdasarkan tingkat loyalitas menggunakan metode *Gaussian Mixture Model* (GMM) dengan pendekatan analisis RFM (*Recency, Frequency, Monetary*). Data transaksi pelanggan dijadikan objek penelitian untuk memahami pola perilaku pelanggan. Tujuan akhirnya adalah membantu Perusahaan mengidentifikasi kelompok pelanggan dengan lebih jelas, sehingga dapat merancang strategi pemasaran yang tepat sasaran.

Dalam analisis ini, *Recency* mengukur seberapa lama waktu sejak transaksi terakhir dilakukan, *Frequency* menunjukkan seberapa sering pelanggan bertransaksi, dan *Monetary* menggambarkan jumlah total uang yang telah dibelanjakan oleh pelanggan. Berdasarkan evaluasi, konfigurasi dengan 5 klaster pada data Primer menggunakan tipe kovarians spherical memberikan hasil terbaik, dengan nilai *silhouette score* tertinggi sebesar 0.8528. Hal ini menunjukkan bahwa pembagian klaster sudah optimal.

Sedangkan konfigurasi dengan 11 klaster pada data sekunder menggunakan tipe kovarians spherical memberikan hasil terbaik, dengan nilai *silhouette score* tertinggi sebesar 0.8101. Hal ini menunjukkan bahwa pembagian klaster sudah optimal.

Pelanggan dalam 5 klaster pada data primer dan 11 klaster pada data sekunder ini dikelompokkan sesuai karakteristik pelanggan. Klaster dengan nilai *Recency*

rendah berisi pelanggan yang baru saja melakukan transaksi, klaster dengan *Frequency* tertinggi mencakup pelanggan sering bertansaksi, sedangkan klaster dengan *Monetary* tertinggi mencerminkan pelanggan dengan kontribusi finansial terbesar. Hasil penelitian ini membuktikan bahwa kombinasi metode GMM dan analisis RFM sangat efektif untuk segmentasi pelanggan, sehingga dapat mendukung perusahaan dalam meningkatkan loyalitas pelanggan dan Menyusun strategi bisnis yang lebih efisien.

## 5.2. Saran

Penelitian selanjutnya dapat menggunakan metode seperti BIC dan AIC untuk menentukan jumlah klaster dalam model GMM. Selain itu, optimasi *hyperparameter*, seperti *covariance*, dapat dilakukan dengan algoritma seperti PSO (*Particle Swarm Optimization*) atau GA (*Genetic Algorithm*).

## DAFTAR PUSTAKA

- Adhinata, F. D., Rakhmadani, D. P., & Segara, A. J. T. (2021). Pengenalan Jenis Kelamin Manusia Berbasis Suara Menggunakan MFCC dan GMM. *Journal of Dinda: Data Science, Information Technology, and Data Analytics*, 1(1), 28–33.
- Agus, A., Mujannah, S., & Fianto, A. Y. A. (2024). Pengaruh Kualitas Pelayanan, Citra Merek dan Kualitas Produk Terhadap Loyalitas Konsumen Melalui Kepuasan Konsumen. *Transformasi Manageria: Journal of Islamic Education Management*, 4(2), 533–549.
- Aini, D. N., Afdal, M., Novita, R., & Mustakim, M. (2024). Implementasi Algoritma Fuzzy C-Means menggunakan Model LRFM untuk Mendukung Strategi Pengelolaan Pelanggan. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 8(3), 1228–1239.
- Ambarsari, E. W., Dwitiyanti, N., Selvia, N., Cholifah, W. N., & Mardika, P. D. (2023). Comparison Approaches of the Fuzzy C-Means and Gaussian Mixture Model in Clustering the Welfare of the Indonesian People. *KnE Social Sciences*, 16–22.
- Dahlan, J., Supriaddin, N., & others. (2022). Mediasi Kepuasan Pelanggan Pada Pengaruh Relationship Marketing Dan Gaya Hidup Terhadap Loyalitas Pelanggan: Studi Pelanggan Kopi Infinite Coffee Shop. *Eqien-Jurnal Ekonomi Dan Bisnis*, 11(03), 629–653.
- Darmawan, N. H. (2022). Segmentasi Loyalitas Pelanggan dengan Model RFM Menggunakan Algoritma Clarans. *Prosiding Sains Nasional Dan Teknologi*, 12(1), 362–368.

- Faidah, D. Y., Hudzaifa, A. M., Theresia, N., & Widianoro, C. E. (2024). OPTIMALISASI STRATEGI PENGELOMPOKKAN POTENSI PADI SEBAGAI SOLUSI EFEKTIF KELANGKAAN BERAS DI JAWA BARAT. *Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, 5(1), 529–537.
- Febriani, A., & Putri, S. A. (2020). Segmentasi Konsumen Berdasarkan Model Recency, Frequency, Monetary dengan Metode K-Means. *JIEMS (Journal of Industrial Engineering and Management Systems)*, 13(2).
- Firmansah, R. Y., Irawan, J. D., & Vendyansyah, N. (2021). Analisis Rfm (Recency, Frequency and Monetary) Produk Menggunakan Metode K-Means. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 5(1), 334–341.
- Gultom, D. K., Arif, M., & Fahmi, M. (2020). Determinasi kepuasan pelanggan terhadap loyalitas pelanggan melalui kepercayaan. *Maneggio: Jurnal Ilmiah Magister Manajemen*, 3(2), 171–180.
- Khaerunnisa, A. (2022). Analisis Tingkat Kelulusan Mahasiswa di Unisba dengan menggunakan Algoritma K-Means Clustering. *J. Ris. Mat*, 67–76.
- Lubis, A. H., Utami, W. R., & Lubis, J. H. (2023). Implementation of k-means clustering for the job provision in urban village. *Jurnal Matematika Dan Ilmu Pengetahuan Alam LLDikti Wilayah 1 (JUMPA)*, 3(1), 21–31.  
<https://doi.org/10.54076/jumpa.v3i1.312>
- Mahmood, H., Mehmood, T., & Al-Essa, L. A. (2023). Optimizing Clustering Algorithms for Anti-Microbial Evaluation Data: A Majority Score-Based Evaluation of K-Means, Gaussian Mixture Model, and Multivariate T-Distribution Mixtures. *IEEE Access*, 11, 79793–79800.

- Nahjan, M. R., Heryana, N., & Voutama, A. (2023). Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 101–104.
- Nurhalizah, R. S., Ardianto, R., & Purwono, P. (2024). Analisis Supervised dan Unsupervised Learning pada Machine Learning: Systematic Literature Review. *Jurnal Ilmu Komputer Dan Informatika*, 4(1), 61–72.
- Oyewole, G. J., & Thopil, G. A. (2023). Data clustering: application and trends. In *Artificial Intelligence Review* (Vol. 56, Issue 7). Springer Netherlands. <https://doi.org/10.1007/s10462-022-10325-y>
- Pailan, Y. O., Chrisnanto, Y. H., & Hadianna, A. I. (2020). Segmentasi Loyalitas Pelanggan Berbasis RFM (Recency, Frequency, Monetary) Menggunakan K-Means pada PD. Persada Ikan. *Prosiding SISFOTEK*, 4(1), 167–171.
- Patel, E., & Kushwaha, D. S. (2020). Clustering cloud workloads: K-means vs gaussian mixture model. *Procedia Computer Science*, 171, 158–167.
- Pramudita, A. S. (2020). Pengukuran Kualitas Pelayanan dan Kepuasan Pelanggan Terhadap Loyalitas Pelanggan pada Industri Logistik. *Jurnal Logistik Bisnis*, 10(1), 15–21.
- 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.
- Prastiwi, H., Pricilia, J., & Rasywir, E. (2022). Implementasi Data Mining Untuk Menentukn Persediaan Stok Barang Di Mini Market Menggunakan Metode K-Means Clustering. *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)*, 2(1), 141–148.

- Pratondo, K., Zaid, Z., Bidul, S., & Aisyah, S. (2024). Pengaruh Kualitas Pelayanan terhadap Loyalitas Pelanggan melalui Citra Perusahaan (Studi Kasus pada Bisnis Ritel TokoMu). *Innovative: Journal Of Social Science Research*, 4(2), 3675–3686.
- Purba, N., Poningsih, P., & Tambunan, H. S. (2021). Penerapan Algoritma K-Means Clustering Pada Penyebaran Penyakit Infeksi Saluran Pernapasan Akut (ISPA) di Provinsi Riau. *Journal of Information System Research (JOSH)*, 2(3), 220–226.
- Rahmattullah, R., Indwiarti Indwiarti, & Rohmawati, A. A. (2023). Clustering Harga Rumah: Perbandingan Model K-Means dan Gaussian Mixture Model. *E-Proceeding Of Engineering*, 10(3), 3441–3449.
- Samara, A., & Susanti, M. (2023). Pengaruh Kemudahan Penggunaan, Pengalaman Pengguna Dan Kepuasan Pelanggan Terhadap Loyalitas Pelanggan Pada Penggunaan Aplikasi Dompot Digital (E-Wallet) di Kalangan Mahasiswa Universitas Buddhi Dharma. *Jurnal Riset Akuntansi*, 1(2), 249–260.
- Sekaran, K., Chandana, P., Krishna, N. M., & Kadry, S. (2020). Deep learning convolutional neural network (CNN) With Gaussian mixture model for predicting pancreatic cancer. *Multimedia Tools and Applications*, 79(15), 10233–10247.
- Setiawan, K. E., & Kurniawan, A. (2023). Pengelompokan Rumah Sakit di Jakarta Menggunakan Model DBSCAN, Gaussian Mixture, dan Hierarchical Clustering. *Jurnal Informatika Terpadu*, 9(2), 149–156.  
<https://doi.org/10.54914/jit.v9i2.995>

- Sidabutar, J. (2020). Penerapan Metode Discrete Wavelet Transform (DWT) dan Gaussian Mixture Model (GMM) Sebagai Pengenal Penutur. *Jurnal Teknik Informatika UNIKA Santo Thomas*, 1–12.
- Trilaksono, B. A., & Handayani, N. (2020). Analisis Value For Money Dan Akuntabilitas Dalam Meningkatkan Pelayanan Publik. *Jurnal Ilmu Dan Riset Akuntansi (JIRA)*, 9(4).
- TU, W., DANG, U. J., & SUBEDI, S. (2024). Change point detection via Gaussian mixture model. *Journal of Statistical Research*, 58(1), 197–219.
- Wahidah, Z., & Utari, D. T. (2023). Comparison Of K-means And Gaussian Mixture Model In Profiling Areas By Poverty Indicators. *Barekeng: Jurnal Ilmu Matematika Dan Terapan*, 17(2), 717–726.
- Wardani, N. W., Hartono, E., & others. (2020). Identifikasi Karakter Pelanggan Retail Berpotensi Churn Dengan Metoda Recency, Frequency Dan Monetary. *Jurnal Teknologi Informasi Dan Komputer*, 6(3).
- Widiyanto, A. T., & Witanti, A. (2021). Segmentasi Pelanggan Berdasarkan Analisis RFM Menggunakan Algoritma K-Means Sebagai Dasar Strategi Pemasaran (Studi Kasus PT Coversuper Indonesia Global). *KONSTELASI: Konvergensi Teknologi Dan Sistem Informasi*, 1(1), 204–215.