SEGMENTASI *CUSTOMER LIFETIME VALUE* MENGGUNAKAN METODE *HUBUNGAN LENGTH RECENCY FREQUENCY MONETARY* DAN ALGORITMA *CLUSTERING EXPECTATION MAXIMIZATION*

SKRIPSI

Oleh:
KUKUH RAHMADANI
NIM. 17650051

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2024

SEGMENTASI *CUSTOMER LIFETIME VALUE* MENGGUNAKAN METODE HUBUNGAN *LENGTH RECENCY FREQUENCY MONETARY* DAN ALGORITMA *CLUSTERING EXPECTATION MAXIMIZATION*

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh:
KUKUH RAHMADANI
NIM. 17650051

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2024

HALAMAN PERTESETUJUAN

SEGMENTASI *CUSTOMER LIFETIME VALUE* MENGGUNAKAN METODE HUBUNGAN *LENGTH RECENCY FREQUENCY MONETARY* DAN ALGORITMA *CLUSTERING EXPECTATION MAXIMIZATION*

SKRIPSI

Oleh:
KUKUH RAHMADANI
NIM. 17650051

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 21 Juni 2024

Pembimbing I,	Pembimbing II,
Syahiduz Zaman, M.Kom	Dr. M. Imamudin.Lc. M.A
NIP. 19700502 200501 1 005	NIP. 19740602 200901 1 010

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Fachrul Kurniawan, M.MT, IPM
NIP. 19771020 200912 1 001

HALAMAN PENGESAHAN

SEGMENTASI *CUSTOMER LIFETIME VALUE* MENGGUNAKAN METODE HUBUNGAN *LENGTH RECENCY FREQUENCY MONETARY* DAN ALGORITMA *CLUSTERING EXPECTATION MAXIMIZATION*

SKRIPSI

Oleh:
KUKUH RAHMADANI
NIM. 17650051

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 21 Juni 2024

Susunan Dewan Penguji

Ketua Penguji	: Prof. Dr. Suhartono, S.Si., M.Kom NIP. 19680519 200312 1 001	( )
Anggota Penguji I	: Dr. M. Ainul Yaqin, M.Kom NIP. 19761013 200604 1 004	( )
Anggota Penguji II	: Syahiduz Zaman, M.Kom NIP. 19700502 200501 1 005	( )
Anggota Penguji III	: Dr. M. Imamudin.Lc. M.A NIP. 19740602 200901 1 010	( )

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Dr. Fachrul Kurniawan, M.MT, IPM
NIP. 19771020 200912 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama	: Kukuh Rahmadani
NIM	: 17650051
Fakultas	: Sains dan Teknologi
Program Studi	: Teknik Informatika
Judul Skripsi	: Segmentasi *Customer Lifetime Value* Menggunakan Metode Hubungan *Length Recency Frequency Monetary* Dan Algoritma *Clustering Expectation Maximization*

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 24 Juni 2024
Yang membuat pernyataan,

Kukuh Rahmadani
NIM. 17650051

MOTTO

“Apabila hal baik tak ada satupun, tunduk lihat ke bawah”

HALAMAN PERSEMBAHAN

Saya persembahkan karya ini kepada:

Ayah saya,

Putut Sarwono

Yang telah mendukung dan menyemangati saya hingga sampai titik ini

Ibu saya,

Sri Ani

Yang telah mendukung dan menyemangati saya hingga sampai titik ini

Teman-teman seperjuangan,

Teknik Informatika Angkatan 2017

Semoga kita semua selalu diberi kemudahan oleh Allah SWT

KATA PENGANTAR

Assalamualaikum Warahmatullahi Wabarakatuh.

Segala puji hanya milik Allah Subhanahu Wa Ta'ala atas segala nikmat dan kasih sayang-Nya yang telah memudahkan penulis untuk menyelesaikan skripsi yang berjudul "Segmentasi *Customer Lifetime Value* Menggunakan Metode Hubungan *Length Recency Frequency Monetary* Dan Algoritma *Clustering Expectation Maximization*". Semoga shalawat dan salam senantiasa terlimpah kepada Nabi Muhammad Sallalahu 'Alaihi wa Sallam. Dan semoga kita semua mendapat syafaatnya di hari kiamat nanti, Aamiin.

Penulis mengucapkan rasa syukur dan terima kasih yang tak terhingga kepada semua pihak-pihak yang selalu memberikan bantuan dan motivasi kepada penulis untuk dapat menyelesaikan skripsi ini. Ucapan ini penulis sampaikan kepada:

1. Prof. Dr. H. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Prof. Dr. Hj. Sri Hariani, M.Si., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Fachrul Kurniawan, M.MT., IPM, selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Syahiduz Zaman, M.Kom selaku dosen pembimbing I dan Dr. M. Imamudin.Lc. M.A selaku dosen pembimbing II yang telah memberikan bantuan dan arahan kepada penulis, sehingga bisa menuntaskan skripsi ini.
5. Prof. Dr. Suhartono, S.Si., M.Kom selaku dosen penguji I dan Dr. M. Ainul Yaqin, M.Kom selaku dosen penguji II yang telah menguji serta

memberikan masukan sehingga penulis dapat menuntaskan skripsi dengan baik.

6. Segenap Dosen, Admin, Laboran dan Mahasiswa Program Studi Teknik Informatika yang telah mencurahkan ilmunya kepada penulis selama kuliah.
7. Kedua orang tua penulis, Bapak Putut Sarwono dan Ibu Sri Ani yang selalu memberi dukungan dan doa untuk penulis dalam menyelesaikan skripsi.
8. Penulis sendiri, yang sudah berusaha sampai pada titik ini dan mampu menyelesaikan skripsi.
9. Teman-teman Teknik Informatika angkatan 2017, yang sudah memberikan energi positif dalam menyelesaikan skripsi.
10. Adelia Nur Sabrina yang terlibat dan membantu secara langsung ataupun tidak langsung dalam menyelesaikan skripsi.

Penulis sepenuhnya menyadari bahwa masih terdapat kekurangan dalam penyusunan skripsi ini. Penulis mengharapkan usulan, saran, dan kritik yang bersifat membangun sebagai upaya tindak lanjut dalam penelitian di skripsi ini. Penulis juga berharap terdapat manfaat yang bisa diambil dari skripsi penulis.

Wassalamualaikum Warahmatullahi Wabarakatuh.

Malang, 24 Juni 2024

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

ABSTRAK

Rahmadani, Kukuh. 2024. **Segmentasi Customer Lifetime Value Menggunakan Metode Hubungan Length Recency Frequency Monetary Dan Algoritma Clustering Expectation Maximization.** Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Syahiduz Zaman, M.Kom. (II) Dr. M. Imamudin.Lc. M.A.

Kata kunci*: *Data Minning, Expectation Maximization, Customer Lifetime Value

Pelanggan merupakan aset penting bagi setiap perusahaan. Memahami nilai pelanggan dan melakukan segmentasi pelanggan yang efektif menjadi kunci untuk meningkatkan profitabilitas dan loyalitas pelanggan. Penelitian ini bertujuan untuk melakukan segmentasi *customer lifetime value* bernilai tinggi dengan menggunakan metode hubungan *length, recency, frequency, monetary* dan algoritma *expectation maximization clustering*. Variabel *length, recency, frequency* dan *monetary* digunakan untuk mengukur nilai pelanggan berdasarkan durasi hubungan l*ength*, frekuensi pembelian r*ecency*, frekuensi transaksi f*requency*, dan nilai transaksi m*onetary*. Algoritma *expectation maximization* kemudian digunakan untuk mengelompokkan pelanggan ke dalam segmen-segmen yang homogen berdasarkan nilai *customer lifetime value* mereka. Dengan 7988 data pelanggan dan 4 variabel yang sudah dijelaskan, yaitu *length, recency, frequency* dan *monetary*. Penelitian ini mengelompokkan menjadi 4 *cluster* yang ditentukan dengan menggunakan metode *silhouette score*. Pengclusteran disimulasikan menggunakan framework flask dengan bahasa pemrograman python 3. Dari hasil simulasi diperoleh cluster 0 dengan nilai loyal terdapat 1135 pelanggan, cluster 1 dengan nilai potential terdapat 2873, cluster 2 dengan nilai general & low value terdapat 3149 pelanggan serta cluster 3 dengan nilai important terdapat 731 pelanggan.

ABSTRACT

Rahmadani, Kukuh. 2024. **Segmentasi Customer Lifetime Value Menggunakan Metode Hubungan Length Recency Frequency Monetary Dan Algoritma Clustering Expectation Maximization.** Thesis. Informatics Engineering Study Program, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University Malang. Supervisor: (I) Syahiduz Zaman, M.Kom. (II) Dr. M. Imamudin.Lc. M.A.

Keywords**:** *Data Minning, Expectation Maximization, Customer Lifetime Value*

Customers are crucial assets for every company. Understanding customer value and effectively segmenting customers is key to enhancing profitability and customer loyalty. This study aims to segment high-value customer lifetime value using the length, recency, frequency, monetary relationship method, and expectation maximization clustering algorithm. Variables such as length (relationship duration), recency (purchase frequency), frequency (transaction frequency), and monetary (transaction value) are used to measure customer value. The expectation maximization algorithm is then employed to cluster customers into homogeneous segments based on their customer lifetime value. With 7988 customer data and the aforementioned 4 variables, namely length, recency, frequency, and monetary, the study identifies 4 clusters determined using the silhouette score method. Clustering was simulated using the Flask framework with Python 3 programming language. Results from the simulation reveal Cluster 0 with 1135 customers characterized as loyal, Cluster 1 with 2873 customers identified as potential, Cluster 2 with 3149 customers categorized as general & low value, and Cluster 3 with 731 customers identified as important.

مستخلص البحث

رحمداني، كوكوه. ٢٠٢٤. تقسيم قيمة حياة العميل باستخدام طريقة طول العلاقة، التردد، النقدية وخوارزمية التكتل التوقعي. رسالة .ماجستير. برنامج دراسة هندسة المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج .المشرف: (الأول) شهيدوز زمان، ماجستير في علوم الحاسوب. (الثاني) الدكتور م. إمامودين. ليسانس الآداب. ماجستير الآداب
الكلمات الرئيسية :تنقيب البيانات، توقعات التكامل

وولاء العملاء. تهدف هذه الدراسة إلى تقسيم قيمة عمر العميل ذات القيمة العالية باستخدام أسلوب العلاقة بين الطول والحداثة والتردد والقيمة المالية وخوارزمية تجميع التوقعات القصوى. يتم استخدام متغيرات مثل الطول (مدة العلاقة) والحداثة (تواتر الشراء) والتردد (تواتر المعاملة) والمالية (قيمة المعاملة) لقياس قيمة العميل. ثم يتم استخدام خوارزمية التوقعات القصوى لتجميع العملاء في قطاعات متجانسة بناءً على قيمة عمر العميل لديهم. باستخدام بيانات ٧٩٨٨ عميلًا والمتغيرات الأربعة المذكورة أعلاه (Silhouette وهي الطول والحداثة والتردد والمالية ، تحدد الدراسة ٤ مجموعات يتم تحديدها باستخدام طريقة درجات الصورة الظلية ، ٣. تكشف نتائج المحاكاة عن المجموعة ٠ التي Python مع لغة البرمجة Flask تمت محاكاة التجميع باستخدام إطار .(score تضم ١١٣٥ عميلًا يتميزون بالولاء ، والمجموعة ١ التي تضم ٢٨٧٣ عميلًا تم تحديدهم كعملاء محتملين ، والمجموعة ٢ التي تضم .٣١٤٩ عميلًا تم تصنيفهم كعملاء عامين وقليلة القيمة ، والمجموعة ٣ التي تضم ٧٣١ عميلًا تم تحديدهم كعملاء مهمين

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pelanggan adalah nyawa bagi setiap perusahaan, perusahaan tidak akan bisa berjalan jika tidak memiliki seorang pelanggan. Pelanggan yang memiliki kepercayaan yang tinggi terhadap perusahaan cenderung akan memilih produk dari perusahaan tersebut saja. Kepercayaan tersebut bisa timbul jika perusahaan mampu memberikan pelayanan ataupun memiliki kualitas produk yang baik dan memiliki harga yang terjangkau. Kualitas pelayanan dapat mempengaruhi seorang pelanggan untuk membeli sebuah produk tersebut ataupun tidak.

Di era globalisasi ini, pelanggan tidak hanya berhubungan dengan perusahaan tetapi juga berhubungan dengan pelanggan lainnya. Dalam sosial media pelanggan dapat menuliskan kritik baik kritik yang positif maupun yang negatif. Jika ulasan yang negatif tersebar lewat media sosial maka akan mempengaruhi oleh pelanggan maupun calon pelanggan. Hal tersebut tentunya sangat merugikan perusahaan. Dahulu ekuitas pelanggan dipandang sebagai aset yang tidak berwujud yang sulit diidentifikasi, tetapi seiring perkembangan teknologi, nilai ekuitas pelanggan dapat diketahui dengan tepat (Pratomo et al., 2019). Profitabilitas pelanggan dan hubungan bisnis antara perusahaan memiliki dampak terbesar pada kinerja keuangan perusahaan manapun. Kehilangan pelanggan akan berdampak pada arus kas masuk perusahaan yang dapat menyebabkan masalah arus kas jangka pendek atau bahkan perusahaan bisa bangkrut (Čermák, 2015) .

Dari permasalahan yang melatarbelakangi, maka perlu dibuatkan sebuah sistem segmentasi pelanggan untuk mengetahui pelanggan mana yang loyal terhadap produk perusahaan dan yang tidak. Untuk itu perusahaan dapat mengetahui mana pelanggan yang harus dipertahankan dan apa yang harus ditingkatkan dalam hal pelayanan maupun dalam hal produk ataupun jasa.

Perusahaan perlu mempelajari konsumen guna mengetahui dan memahami dalam berbagai aspek yang ada pada konsumen. Dalam mempelajari konsumen sangatlah penting, karena perusahaan dapat menjalankan bisnisnya serta mendapatkan nilai pelanggan atau *Customer Lifetime Value* (CLV) bagi perusahaan tersebut. Dengan mendapatkan nilai pelanggan maka perusahaan membuat keputusan yang strategis guna meningkatkan loyalitas pelanggan serta bermanfaat untuk memformulasikan strategi promosi (Ditendra et al., 2020).

Al-Qur'an sebagai sumber hukum dan petunjuk hidup bagi umat Islam mengandung banyak sekali ajaran yang relevan dengan berbagai aspek kehidupan, termasuk dalam bidang ekonomi dan bisnis. Salah satu ayat yang menekankan pentingnya pencatatan dan pengelolaan transaksi adalah Surah Al-Baqarah ayat 282 yang berbunyi:

يٰٓاَيُّهَا الَّذِيْنَ اٰمَنُوْٓا اِذَا تَدَايَنْتُمْ بِدَيْنٍ اِلٰٓى اَجَلٍ مُّسَمًّى فَاكْتُبُوْهُۗ

"Hai orang-orang yang beriman, apabila kamu bermuamalah tidak secara tunai untuk waktu yang ditentukan, hendaklah kamu menuliskannya. Dan hendaklah seorang penulis di antara kamu menuliskannya dengan benar..." (Al-Baqarah: 282).

Ayat tersebut mengajarkan tentang pentingnya kejelasan dan ketelitian dalam setiap transaksi, yang dapat diartikan sebagai prinsip pengelolaan data dan informasi secara akurat dan sistematis. Dalam konteks bisnis modern, pengelolaan data pelanggan menjadi sangat penting untuk memahami perilaku konsumen dan meningkatkan kualitas pelayanan.

Peneliti akan membuat sistem segmentasi pelanggan dengan model LRFM (*Length, Recency, Frequency dan Monetary*) dan metode *Expectation Maximization Clustering* (EMC) dan akan menghasilkan *Customer Lifetime Value* (CLV). Model LRFM digunakan untuk memperoleh nilai segmentasi sementara terhadap perilaku pelanggan sedangkan *Expectation Maximization Clustering* digunakan untuk segmentasi hasil sementara dari model LRFM untuk mendapatkan nilai CLV.

1.2 Rumusan Masalah

Bagaimana implementasi segmentasi *Customer Lifetime Value* bernilai tinggi menggunakan metode hubungan LRFM dan algoritma *Expectation Maximization Clustering* ?

1.3 Tujuan Penelitian

Melakukan segmentasi *Customer Lifetime Value* dengan METODE hubungan LRFM dan algoritma *Expectation Maximization Clustering*.

1.4 Batasan Masalah

1. Algoritma yang digunakan untuk segmentasi pelanggan adalah *Expectation Maximization Clustering* dengan model LRFM.

2. Data merupakan data sekunder, didapatkan dari website www.kaggle.com sebagai pola model untuk dilakukan segmentasi pelanggan.

1.5 Manfaat Penelitian

1. Untuk mengetahui segmentasi pelanggan dengan model LRFM dan algoritma *Expectation Maximization Clustering* untuk segmentasi *Customer Lifetime Value*.
2. Mendapatkan tingkat akurasi algoritma segmentasi pelanggan menggunakan model LRFM dan metode *Expectation Maximization Clustering*.

BAB II

STUDI PUSTAKA

Pada bab ini berisi tinjauan pustaka yang membahas beberapa penelitian terkait, landasan teori, dan metode yang digunakan penulis sebagai acuan mengerjakan penelitian tugas akhir ini.

2.1 Penelitian Terkait

Penelitian clustering *Customer Lifetime Value* (CLV) menggunakan algoritma K-Means dan model LRFM (*Length*, *Recency*, *Frekuensi*, *Monetary*). Nilai Seumur Hidup Pelanggan, yang merupakan faktor penting dalam analisis pelanggan, mengukur nilai potensial yang dapat ditawarkan klien kepada bisnis dalam jangka panjang. Dengan memperhitungkan variabel-variabel seperti *length* (lama menjadi pelanggan), *recency* (waktu transaksi klien terkini), *frequency* (frekuensi transaksi), dan *monetary* (jumlah transaksi), model LRFM telah memantapkan dirinya sebagai teknik umum untuk memperkirakan CLV.

Dalam penelitian ini, penulis berhasil mengelompokkan pelanggan berdasarkan karakteristik transaksionalnya dengan menggunakan model LRFM dan algoritma K-Means. Langkah-langkah analisis data, konstruksi model LRFM, dan penerapan algoritma K-Means telah menghasilkan pemahaman yang lebih mendalam tentang perilaku konsumen dan membantu bisnis dalam mengembangkan strategi pemasaran dan retensi klien yang lebih efektif. Temuan penelitian ini memberikan informasi penting tentang bagaimana meningkatkan

analisis CLV dengan menggunakan model LRFM dan algoritma K-Means untuk segmentasi pelanggan yang lebih baik (Monalisa, 2018).

Pada penelitian ini, metode *Expectation Maximization* (EM) diterapkan untuk meramalkan penjualan susu murni di PT. Sewu Primatama Indonesia di Lampung Tengah. Model perkiraan penjualan susu murni dioptimalkan menggunakan pendekatan EM, yang memperhitungkan sejumlah faktor yang mempengaruhi penjualan, termasuk waktu dalam setahun, harga, promosi saat ini, dan informasi penjualan sebelumnya. Agar bisnis dapat merencanakan produksi dan persediaan inventaris dengan lebih efektif, penelitian ini memanfaatkan data penjualan yang telah dikumpulkan dari perusahaan dan difokuskan pada peningkatan prediksi penjualan.

Hasil penelitian menunjukkan bahwa jika dibandingkan dengan teknik prediksi konvensional, algoritma *Expectation Maximization* dapat menghasilkan proyeksi penjualan susu murni yang lebih akurat. Akibatnya, PT. Sewu Primatama Indonesia, Lampung Tengah, mampu mengelola inventaris dengan lebih efektif, mengurangi kerugian akibat kekurangan stok, dan meningkatkan kepuasan pelanggan dengan menyediakan lebih banyak produk. Informasi yang diperoleh dari penelitian ini juga dapat digunakan untuk membuat rencana pemasaran masa depan yang lebih sukses dengan memberikan wawasan penting mengenai variabel-variabel yang mempengaruhi penjualan susu murni di wilayah tersebut (Marisa Efendi et al., 2022).

Penelitian untuk menentukan jumlah cluster pada data dengan mengambil variabel bobot menjadi pertimbangan. Pendekatan *Gaussian Mixture Model*

(GMM) dan *K-Means*, yang tidak mampu menangani fluktuasi bobot dalam data, direkomendasikan sebagai alternatif metodologi ini.

Untuk setiap variabel dalam data dalam penelitian ini, WGM menghasilkan rata-rata terbobot dan nilai kovarians. *Bayesian Information Criterion* (BIC) yang diperoleh dengan menggunakan bobot variabel kemudian digunakan untuk menghitung jumlah *cluster*. Temuan eksperimental pada data buatan dan nyata menunjukkan bahwa WGM lebih tepat daripada GMM dan K-Means dalam memperkirakan jumlah cluster, terutama untuk data dengan bobot variabel yang tinggi.

Bayesian Information Criterion (BIC) digunakan oleh WGM untuk menghitung jumlah klaster yang ideal setelah menerapkan pembobotan ke variabel untuk membuat nilai rata-rata dan kovarian tertimbang. Hasil pengujian menunjukkan bahwa WGM, khususnya pada data dengan bobot variabel yang substansial, memberikan jumlah *cluster* yang lebih tepat dibandingkan GMM dan *K-Means*. Jadi, algoritme WGM mungkin merupakan pengganti yang masuk akal untuk mengetahui berapa banyak cluster yang ada dalam data dengan bobot yang berubah. (Chakraborty & Das, 2018).

Penerapan dan pengujian Algoritma Clustering Expectation-Maximization (EM) dalam kaitannya dengan Data Tugas Akhir Telkom University dibahas pada penelitian sebelumnya ini. Algoritma EM sekarang banyak digunakan dalam analisis data, khususnya dalam pekerjaan pengelompokan data yang sulit. Tujuan dari penelitian ini adalah untuk menyelidiki efektivitas penggunaan algoritma EM terhadap data yang dihasilkan oleh mahasiswa untuk proyek batu penjuru mereka.

Penelitian ini akan menguji efektivitas algoritma EM dalam mengklasifikasikan data dengan menggunakan dataset yang berasal dari tugas akhir mahasiswa. Hal ini juga akan menilai keuntungan dan kerugian penerapan algoritma ini di ruang kelas.

Hasil penelitian ini memiliki konsekuensi yang signifikan untuk meningkatkan pengetahuan tentang penerapan algoritma EM di perguruan tinggi, khususnya di Telkom University. Selain itu, hasil penelitian ini dapat membantu akademisi universitas dan pengambil keputusan dalam mengoptimalkan penggunaan algoritma EM untuk analisis data tugas akhir mahasiswa. Hasilnya, bab penelitian ini secara signifikan meningkatkan efektivitas dan efisiensi klasifikasi data akademik dan mendukung pengambilan keputusan berdasarkan data di lingkungan pendidikan tinggi (Sirait et al., n.d.).

Penelitian selanjutnya berfokus pada mengembangkan sistem berbasis pengelompokan *G-Means* untuk mengumpulkan data untuk kegiatan ilmiah. Pengguna diantisipasi untuk mendapatkan keuntungan dari bantuan penelitian ini dalam memperoleh informasi ilmiah saat ini dan dapat diandalkan.

Penulis penelitian ini menggunakan kumpulan data yang terdiri dari makalah akademik dari konferensi global tentang ilmu komputer dan teknologi informasi. Dataset dibagi menjadi jumlah cluster yang ideal menggunakan pendekatan *clustering G-Means*. Sistem kemudian akan menentukan kata kunci mana yang paling lazim di setiap klaster untuk menyarankan karya ilmiah yang relevan dan terkait. Hasil pengujian menunjukkan bahwa sistem dapat membuat rekomendasi akurat yang sesuai dengan kebutuhan pengguna.

Membangun metode pengumpulan informasi yang andal dan efisien untuk kegiatan ilmiah adalah salah satu manfaat dari penelitian ini. Saat mempartisi kumpulan data ke dalam jumlah grup yang ideal, algoritma pengelompokan *G-Means* dapat mengurangi kesalahan. Selain itu, pendekatan ini memudahkan pengguna untuk menemukan informasi tentang artikel akademis yang relevan dan terkait dengan kata kunci pilihan mereka. Pengembangan sistem pencarian informasi karya ilmiah di masa mendatang diharapkan dapat bermanfaat dari penelitian ini. (Agri Ardyan & Budi Darmawan, 2016).

Tabel 2.1 Novelty Penelitian terkait

No	Input	Metode	Hasil	Limitasi
1	Data transaksi pelanggan (jumlah transaksi, *recency*, *frequency*, dan *monetary*)	Pra Metode: Dunn Index Main Metode: K-Means	Clustering CLV	Studi ini hanya menggunakan empat fitur data, yaitu jumlah transaksi, *recency*, *frequency*, dan *monetary*. Fitur lain, seperti jenis produk yang dibeli, lokasi pelanggan, dan waktu pembelian, juga dapat digunakan untuk meningkatkan akurasi hasil pengelompokan (S. Monalisa., 2018)
2	Data penjualan susu murni PT. Sewu Primatama	Main Metode: Expectation Maximitation	Cluster Model	Studi ini hanya menggunakan data penjualan susu murni PT. Sewu Primatama Indonesia Lampung Tengah dari tahun 2019-2021 untuk melatih model prediksi. Data penjualan susu murni dari tahun-tahun sebelumnya atau dari perusahaan susu murni lainnya dapat digunakan untuk meningkatkan akurasi prediksi (Marisa et al., 2022)
3	Data dengan fitur-fitur yang beragam	Pra Metode: Gaussian Mixture Model Main Metode: K-Means	Cluster Model	Studi ini hanya mengevaluasi kinerja algoritma WGM pada data sintetis dan data nyata dengan jumlah fitur yang relatif sedikit (Chakraborty et al., 2018)
4	Data tugas akhir Universitas Telkom (judul, abstrak, dan kata kunci)	Main Metode: Expectation Maximitation	Cluster Model	Ukuran Dataset terbatas Sirait (R, Darwianto E, Dwi D et al., 2015)
5	Pengunjung objek wisata di Malang Raya	Main Metode: Expectation Maximitation	Cluster Model	Hanya menggunakan satu indikator popularitas dan tidak mempertimbangkan distribusi spasial objek wisata (Atikah N., 2019)
6	Data tugas akhir Universitas Telkom (judul, abstrak, dan kata kunci)	Main Metode: G-Means Clustering	Cluster Model	Penelitian ini hanya menggunakan tiga fitur dokumen, yaitu judul, abstrak, dan kata kunci. Fitur lain, seperti nama penulis, tahun publikasi, dan jurnal tempat diterbitkannya (Sirait et al., n.d., 2019)

Tidak ada dua penelitian yang sama, dan penelitian ini membuka pintu pada bidang yang relatif belum dieksplorasi dalam analisis Customer Lifetime Value (CLV) dengan menggabungkan metode Hubungan LRFM dan Algoritma Clustering Expectation Maximization.

1. Penggabungan Metode yang Berbeda: Penelitian ini memadukan metode Hubungan LRFM (*Length, Recency, Frequency, Monetary*) dan Algoritma *Clustering Expectation-Maximization*. Ini merupakan pendekatan baru yang mengintegrasikan data perilaku pelanggan dengan algoritma clustering yang canggih. Studi sebelumnya mungkin hanya menggunakan satu metode atau algoritma, sementara penelitian Anda menggabungkan keduanya untuk analisis yang lebih komprehensif.
2. Spesifik pada *Customer Lifetime Value*: Penelitian ini berfokus pada segmentasi berdasarkan *Customer Lifetime Value*, yang mencerminkan kontribusi jangka panjang pelanggan terhadap bisnis. Sementara banyak penelitian sebelumnya lebih terfokus pada segmentasi pelanggan berdasarkan perilaku transaksional atau demografis, penelitian Anda menyoroti pentingnya menganalisis segmen pelanggan berdasarkan nilai jangka panjang.
3. Dampak pada Strategi Pendidikan dan Bisnis: Penelitian Anda mungkin menghasilkan temuan yang dapat memengaruhi strategi pendidikan dan bisnis universitas. Hal ini dapat mencakup rekomendasi terkait dengan program akademik, manajemen sumber daya, atau pengambilan keputusan dalam konteks pendidikan tinggi.

2.2 Customer Relationship Management (CRM)

Custumer Relationship Management (CRM) merupakan sebuah filosofi bisnis yang menggambarkan suatu strategi penempatan client sebagai pusat proses, aktivitas dan budaya. Konsep ini telah dikenal dan banyak diterapkan untuk meningkatkan pelayanan di perusahaan (Dyantina et al., 2012). Didalam singkatan CRM tersebut terdapat singkatan relationship (hubungan) yang artinya adalah suatu hubungan terdiri atas serangkaian episode yang terjadi antara dua pihak pada rentang waktu yang tertentu. Didalam hubungan kadang-kadang mengalami pasang surut (akan terjadi evolusi dalam hubungan tersebut). Model yang bisa dikembangkan dalam relationship yakni kepercayaan dan komitmen. Kepercayaan bisa datang ketika kedua belah pihak saling berbagi pengalaman, mereka mulai bisa saling memahami satu dengan yang lainnya. Sedangkan komitmen adalah keyakinan dari salah satu mitra akan pentingnya arti membangun hubungan jangka panjang yang langgeng dengan mitranya. Komitmen akan muncul sebagai buah dari pada kepercayaan (Rosmayani, 2016).

Kerangka komponen CRM diklasifikasikan menjadi tiga yaitu (Dyantina et al., 2012) :

1. Operasional CRM

Operasional CRM dikenal sebagai front office perusahaan. Komponen CRM ini berperan dalam interaksi dengan pelanggan. Operasional CRM mencakup proses otomatisasi yang terintegrasi dari keseluruhan proses bisnis, seperti otomatisasi pemasaran, dan pelayanan. Salah satu penerapan CRM yang termasuk

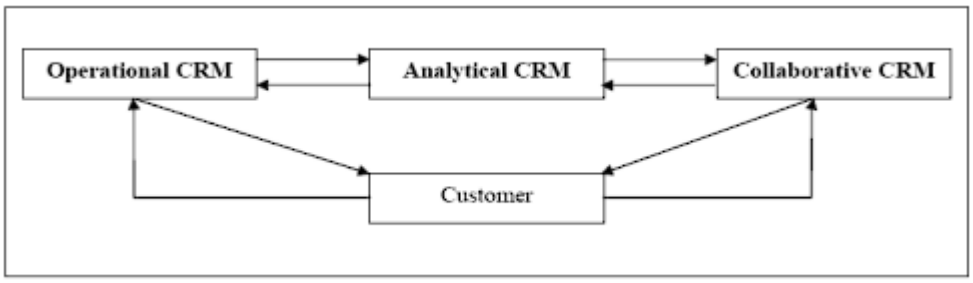
dalam kategori operasional CRM adalah dalam bentuk aplikasi web. Melalui web, suatu perusahaan dapat memberikan pelayanan kepada pelanggan.

2. Analitikal CRM

Analitikal CRM dikenal sebagai back office perusahaan. Komponen CRM ini berperan dalam memahami kebutuhan pelanggan. Analitikal CRM berperan dalam melaksanakan analisis pelanggan dan pasar, seperti analisis trend pasar dan analisis kebutuhan dan perilaku pelanggan. Data yang digunakan pada CRM analitik adalah data yang berasal dari CRM operasional.

3. Collaborative CRM

Komponen kolaborasi CRM meliputi e-mail, personalized publishing, ecommunities, dan sejenisnya yang dirancang untuk interaksi antara pelanggan dengan perusahaan. Tujuan utamanya adalah memberikan nilai tambah dan memperluas loyalitas pelanggan ke pelanggan lain yang masih belum berada di level kesetiaan pelanggan. Collaborative CRM juga mencakup pemahaman atau kesadaran bahwa pelanggan yang setia dapat menjadi magnet bagi pelanggan lain.



Gambar 2.1 Kerangka Customer Relationship Management (CRM)

2.3 Segmentasi Pelanggan

Segmentasi pelanggan mengacu pada pembagian sasaran pasar menjadi kelompok yang dapat dikelola dan layak sesuai dengan karakteristik bersama untuk

mengembangkan strategi bisnis yang efektif dan tepat (Parvaneh et al., 2012). Secara sederhana, pembagian segmentasi ini adalah proses membagi basis pelanggan yang ada menjadi kelompok yang dapat dikelola dan layak berdasarkan karakteristik umum seperti usia, jenis kelamin, loyalitas, frekuensi pembelian, dll. Untuk menargetkan dan mengembangkan strategi pemasaran untuk setiap kelompok sesuai dengan karakteristik itu (Kartika Zahretta Wijaya et al., 2021).

2.4 Peranan Segmentasi dalam Pemasaran

Peranan segmentasi dalam marketing dilakukan dengan membagi beberapa pelanggan yang ada dalam aturan tertentu. Cara untuk melakukan segementasi pelanggan ialah dengan cara melakukan klustering. Tujuan segmentasi pelanggan termasuk membagi pelanggan target menjadi kelompok-kelompok yang lebih kecil yang mencerminkan kesamaan di antara pelanggan di setiap kelompok untuk (Christy et al., 2021) :

1. Mengembangkan hubungan yang lebih baik dengan memahami kebutuhan setiap segmen pelanggan
2. Mengidentifikasi pelanggan yang berharga
3. Mengidentifikasi peluang cross-selling dan up-selling
4. Meningkatkan profitabilitas dengan mengembangkan strategi pemasaran yang lebih efektif untuk setiap segmen

2.5 Strategi Pemasaran

Strategi pemasaran merupakan salah satu dari beberapa cara untuk memenangkan keunggulan dalam bersaing yang berkelanjutan baik dalam perusahaan yang memproduksi barang maupun jasa. Strategi pemasaran dapat

diartikan sebagai bentuk salah satu landasan yang digunakan dalam penyusunan perencanaan perusahaan secara total. Dipandang dari segi luasnya permasalahan yang ada dalam sebuah perusahaan, dengan demikian diperlukan adanya perencanaan teknis yang menyeluruh untuk dijadikan acuan bagi perusahaan dalam menjalankan kegiatannya (Gesta Nabilla & Tuasela, 2021).

Strategi pemasaran merupakan salah satu yang menjadi hal terpenting dalam merintis potensi kemajuan bisnis dalam hal hubungan antara pelanggan dengan perusahaan. Oleh karena itu, untuk membangun hubungan antara pelanggan dan perusahaan agar menjadi lebih baik, diperlukannya berbagai cara manajemen untuk memecahkan permasalahan dan pengambilan keputusan yang bersifat strategis (Moh rusdi, 2019). Sehingga dapat terciptanya loyalitas pelanggan terhadap perusahaan. Dengan terciptanya loyalitas pelanggan maka pelanggan akan tidak ragu untuk mengeluarkan uang mereka untuk bertransaksi.

2.6 Length Recency Frequency Monetary (LRFM)

RFM merupakan model yang banyak digunakan dalam melakukan segmentasi perilaku pelanggan. Model RFM ini terdiri dari Recency, Frequency, dan Monetary (Monalisa, 2018b). Metode ini telah dikembangkan menjadi LRFM dengan mengambil *Length* relasi pelanggan terhadap perusahaan yang bertujuan untuk memecahkan masalah dari model RFM yang mana pada RFM terdapat kesulitan untuk membedakan hubungan jangka panjang dan hubungan jangka pendek pelanggan terhadap perusahaan. Length dari interval relasi pelanggan terhadap perusahaan sangatlah penting untuk menentukan loyalitas dari pelanggan

(Ditendra et al., 2020b). Definisi lebih lengkap dari atribut LRFM dapat dilihat pada tabel 2.2 berikut.

Tabel 2.2 Penjelasan Atribut LRFM

No	Atribut	Definisi
1	*Length*	Merupakan panjangnya interval hubungan antara pelanggan dengan perusahaan dengan melihat transaksi awal dan transaksi terakhir pelanggan pada periode yang telah ditentukan.
2	*Recency*	Merupakan waktu pelanggan melakukan transaksi terakhir pada periode yang telah ditentukan.
3	*Frequency*	Merupakan jumlah dari transaksi pelanggan terhadap perusahaan pada periode yang telah ditentukan.
4	*Monetary*	Merupakan jumlah uang yang dikeluarkan pelanggan terhadap perusahaan yang berasal dari transaksi selama periode yang telah ditentukan.

2.7 Customer Lifetime Value (CLV)

Customer Lifetime Value (CLV) merupakan konsep manajemen dalam hubungan terhadap pelanggan / *Customer Relationship Management* (CRM). Hal tersebut didefinisikan sebagai nilai sekarang dari semua keuntungan keuntungan masa depan yang dapat diperoleh dari pelanggan selama masa hubungan dengan perusahaan. Pemasaran secara langsung merupakan tindakan memperlakukan pelanggan secara berbeda berdasarkan tingkat keuangan mereka dan CLV adalah indikator paling andal dalam pemasaran langsung untuk mengukur profitabilitas pelanggan (Ditendra et al., 2020b). Metode ini digunakan untuk melakukan perhitungan nilai profitabilitas terhadap pelanggan. Setelah melakukan segmentasi terhadap pelanggan, CLV dihitung. Perhitungan nilai CSV dilakukan pada rangking CLV yang telah ditentukan pada setiap segmen pelanggan. Perhitungan CLV dapat dipresentasikan pada persamaan berikut.

$$C^J = W_L C_L^J + W_R C_R^J + W_F C_F^J + W_M C_M^J \quad (2.1)$$

Yang mana:

C^J = Peringkat CLV Pelanggan J

W_L, W_R, W_F, W_M = Bobot yang dihasilkan L,R,Fdan M

$C_L^J, C_R^J, C_F^J, C_M^J$ = Normalisasi L,R,F,M dari Kluster J

2.8 Scaling Min Max

Metode *scaling min max adalah* salah satu pendekatan normalisasi yang sering digunakan dalam analisis data dan pemrosesan data adalah metode penskalaan data dengan *scaling min max*. Metode ini bertujuan untuk menggeser rentang nilai fitur atau variabel ke dalam rentang yang telah ditentukan, seperti dari 0 hingga 1. Untuk memverifikasi bahwa setiap fitur dalam dataset memiliki skala yang seragam dan berkontribusi berbeda, digunakan skala data menggunakan metode *min max*. dibandingkan dalam hal bagaimana analisis dan pengambilan keputusan dilakukan (Gopal et al., n.d.).

Dalam metode *scaling* data dengan *min max,* langkah-langkah yang dilakukan meliputi:

1. Mengidentifikasi fitur yang akan dinormalisasi.
2. Menentukan rentang normalisasi yang diinginkan, misalnya dari 0 hingga 1.
3. Menghitung nilai minimum dan maksimum dari fitur yang akan dinormalisasi.
4. Menggunakan rumus *min max* untuk mengubah setiap nilai dalam fitur menjadi rentang normalisasi. Rumusnya adalah sebagai berikut:

$$x' = (x - min) / (max - min) \quad (2.2)$$

Yang mana:

x' = Nilai yang dinormalisasi

x = Nilai asli data

min = Nilai minimum dari data

max = Nilai maksimum dari data

2.9 Principal Component Analysis

Principal Component Analysis (PCA) disebut juga dengan Analisis Komponen Utama adalah pengurangan dimensi suatu kumpulan data yang terdiri dari sejumlah besar variabel yang saling terkait, dengan mempertahankan sebanyak mungkin variasi yang ada. Hal ini dicapai dengan mentransformasikan data ke sekumpulan variabel baru, komponen utama (PC), yang tidak berkorelasi, dan mana yang berkorelasi diurutkan sedemikian rupa sehingga beberapa yang pertama mempertahankan sebagian besar variasi yang ada di semua variabel aslinya. Secara umum, tujuan utama Principal Component Analysis (PCA) adalah untuk mengurangi kompleksitas hubungan timbal-balik antara sejumlah besar variabel yang diamati ke sejumlah relatif kecil dari kombinasi linearnya, yang disebut sebagai komponen utama. Karena jumlah komponen utama sama banyaknya dengan jumlah variabel dalam data, maka komponen utama disusun sedemikian rupa sehingga komponen utama pertama menyumbang kemungkinan varian terbesar dalam kumpulan data (Jolliffe. IT, 2019).

Analisis Komponen Utama biasanya digunakan untuk:

1. Mengidentifikasi variabel baru yang mendasari data variabel ganda.
2. Mengurangi banyaknya dimensi himpunan variabel yang biasanya terdiri atas variabel yang banyak dan saling berkolerasi dengan mempertahankan sebanyak mungkin keragaman dalam himpunan data tersebut.
3. Menghilangkan variabel-variabel asal yang mempunyai sumbangan informasi yang relatif kecil. Variabel baru yang dimaksud di atas disebut komponen utama yang mempunyai ciri yaitu :
    a. Merupakan kombinasi linier variabel-variabel asal.
    b. Jumlah kuadrat koefisien dalam kombinasi linier tersebut bernilai satu.
    c. Tidak berkorelasi, dan mempunyai ragam berurut dari yang terbesar ke yang terkecil.

2.10 Data Mining

Data *mining* adalah kegiatan mengekstraksi atau menambang pengetahuan dari data yang berukuran/berjumlah besar, informasi inilah yang nantinya sangat berguna untuk pengembangan. Definisi sederhana dari data mining adalah ekstraksi informasi atau pola yang penting atau menarik dari data yang ada di database yang besar. Dalam jurnal ilmiah, data *mining* juga dikenal dengan nama *knowledge discovery in databases* (KDD) (Heri Susanto, 2014).

Data *mining* merupakan konsep penggalian tersembunyi informasi dari *database* dalam jumlah yang besar dan melakukan pendeteksian pola yang menarik secara Himpunan data utuh. Data *mining* merupakan sebuah proses menganalisis data dari perspektif yang berbeda dan merangkainya menjadi informasi yang

mempunyai nilai guna. Tujuan dari data *mining* yaitu untuk menggali informasi yang berguna untuk mengurangi biaya dan meningkatkan pendapatan atau keduanya (Damuri et al., 2021).

2.11 Metode Expectation Maximization (EM)

Algoritma EM yaitu algoritma yang berfungsi untuk menemukan nilai estimasi *maximum likelihood* dari parameter dalam sebuah model probabilistik (Nur Atikah et al., 2021). Kelebihan dari algoritma EM adalah bisa menyelesaikan permasalahan bidang statistik antara lain pendugaan parameter untuk gabungan fungsi-fungsi serta parameter dari data yang tidak lengkap. Dalam algoritma ini, ada dua hal yang digunakan secara bergantian yaitu *E-step* yang digunakan untuk menghitung nilai ekspektasi dari *likelihood* dan *M-step* digunakan untuk menghitung nilai estimasi dari parameter dengan memaksimalkan nilai ekspektasi dari *likelihood* yang ditemukan pada *E-step.*

Metode *Expectation Maximization* (EM) merupakan algoritma segementasi yang bertujuan untuk memperoleh perkiraan hasil yang baik dengan memaksimalkan fungsi kemungkinan. EM juga termasuk algoritma cluster yang menggunakan model perhitungan probabilitas.

Metode ini menggunakan kecerdasan buatan, matematika, dan metode statistik untuk mengekstraksi informasi yang berguna dari data yang ada. Penambangan data digunakan dalam konteks tesis ini sebagai alat analisis untuk menyelidiki hubungan antara variabel yang penting untuk penelitian. Analisis regresi, salah satu pendekatan penambangan data yang paling banyak digunakan, memungkinkan peneliti untuk memahami bagaimana satu variabel memengaruhi

variabel lainnya. Selain itu, algoritma klasifikasi sering digunakan dalam penambangan data untuk mengklasifikasikan data berdasarkan ciri-ciri yang termasuk dalam dataset (Smith, 2020).

BAB III

DESAIN DAN IMPLEMENTASI

3.1 Desain Penelitian

3.1.1 Gambran Umum Sistem

Penelitian ini akan membahas tentang bagaimana melakukan segmentasi *Customer Lifetime Value* (CLV) dengan menggunakan hubungan LRFM (*Length, Recency, Frequency, Monetary*) dengan algoritma *Clustering Expectation Maximization*. Metode hubungan LRFM digunakan untuk menganalisis perilaku pelanggan berdasarkan *length, recency, frequency,* dan *monetary value* dari transaksi yang telah dilakukan oleh setiap pelanggan. Sedangkan algoritma *Clustering Expectation Maximization* digunakan untuk mengelompokkan pelanggan ke dalam segmen karateristik LRFM yang dimiliki.

Customer Lifetime Value dengan metode hubungan LRFM. Panjang (*length*) mencerminkan komitmen pelanggan, kebaruan (*recency*) menunjukkan tingkat keterlibatan saat ini, frekuensi transaksi (*frequency*) menunjukkan intensitas aktivitas pelanggan, dan moneter nilai (*monetary*) menunjukkan kontribusi keuangan pelanggan terhadap perusahaan. Penelitian ini berupaya memperoleh wawasan menyeluruh tentang perilaku dan preferensi konsumen dengan menggabungkan keempat komponen tersebut. Teknik *Clustering Expectation Maximization* juga akan digunakan untuk mengidentifikasi tren dan sifat yang muncul dalam data LRFM. Pelanggan akan dikelola dengan lebih baik dan rencana pemasaran dapat disesuaikan sebagai hasil dari segmentasi pelanggan, yang juga

akan meningkatkan pendapatan perusahaan secara keseluruhan dan retensi pelanggan.

Kombinasi dari metode hubungan LRFM dan algoritma *Clustering Expectation Maximization* adalah metode baru dan menjanjikan untuk analisis *Customer Lifetime Value*. Sementara teknik *Clustering Expectation Maximization* memungkinkan untuk mengelompokkan data secara lebih efektif dan akurat, metode LRFM memungkinkan analisis perilaku pelanggan yang lebih detail dan akurat. Selain membantu bisnis dalam mengidentifikasi segmen pelanggan yang berpotensi menghasilkan nilai CLV yang tinggi, diharapkan penelitian ini juga memungkinkan terciptanya strategi pemasaran yang lebih individual dan efisien yang akan meningkatkan keterlibatan dan kepuasan pelanggan.

3.1.2 Sumber Data

Data yang digunakan dalam penelitian ini adalah data *secondary* yang diambil dari kaggle. Adapun proses pengambilan datanya dilakukan pada bulan Maret 2023. Data kemudian diolah. Informasi yang sudah ada dan telah dikumpulkan oleh pihak lain untuk tujuan yang berbeda tetapi dapat digunakan kembali untuk analisis baru.

Pengambilan data dari website Kaggle dilakukan dengan mengidentifikasi kolom-kolom tertentu yang akan digunakan sebagai atribut dalam analisis segmentasi pelanggan berdasarkan metode *LRFM* :

Table 3.1. Pembagian kolom *dataset* ke model LRFM

No	Model LRFM	Kolom Dataset
1	Length	FFP_DATE, LOAD_TIME
2	Recency	LAST_TO_END
3	Frequency	FLIGHT_COUNT
4	Monetary	SEG_KM_SUM

Kelebihan menggunakan data *secondary* dalam penelitian ini adalah sebagai berikut:

1. Ketersediaan Data

Data sekunder seringkali dapat diakses secara luas dan mudah diperoleh. Hal ini disebabkan bahwa data sudah dikumpulkan untuk tujuan penelitian komersial atau lainnya, menghemat waktu dan uang selama proses pengumpulan data.

2. Historis Data

Riwayat transaksi klien lama sering disertakan dalam data sekunder, memungkinkan analisis perilaku konsumen yang lebih menyeluruh dari waktu ke waktu. Memahami tren dan pola yang dapat berkembang dari waktu ke waktu menjadi lebih mudah dengan melakukan ini.

3. Keberagaman Data

Data sekunder seringkali berasal dari berbagai sumber dan berisi berbagai faktor, seperti data demografis, pola pembelian, preferensi produk, dan sebagainya. Berbagai jenis data ini dapat memberi kami wawasan yang lebih luas dan mendalam tentang bagaimana perilaku pelanggan.

4. Potensi Integrasi Data

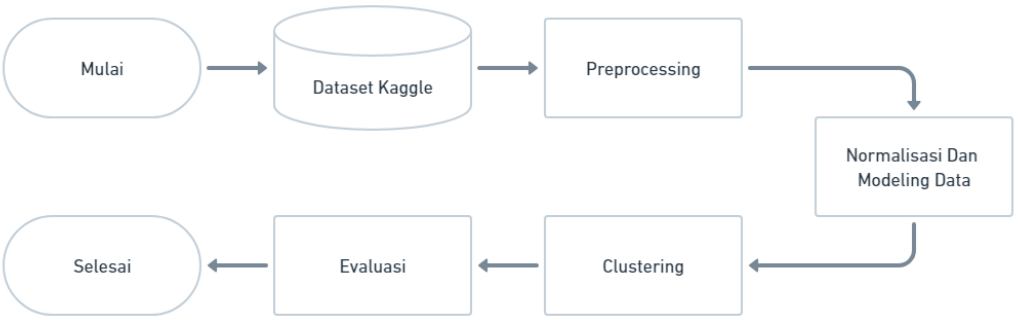
Data internal perusahaan atau data dari sumber lain, seperti data pemasaran, data transaksi, atau data dari media sosial, seringkali dapat digabungkan dengan data sekunder. Keakuratan segmentasi dapat ditingkatkan dengan menggunakan integrasi data ini untuk memberikan gambaran yang lebih lengkap kepada klien.

5. Penghematan Biaya dan Waktu

Peneliti dapat menghemat uang dan tenaga dengan mengandalkan data sekunder daripada melakukan pengumpulan data sendiri. Hal ini memungkinkan untuk lebih berkonsentrasi pada analisis dan interpretasi hasil.

3.2 Perancangan Sistem

Sistem klasifikasi ini dibuat dengan rancangan alur langkah demi langkah, dimulai dengan pengumpulan data sampai diakhiri dengan klasifikasi sehingga dapat dilakukan perhitungan untuk menentukan keakuratan proses pengujian. Sentimen analitik dan pendekatan kategorisasinya terkait dengan setiap langkah dalam alur sistem. Oleh karena itu, penulis membuat perancangan olah dataset yang ditunjukkan pada Gambar 3.1.



Gambar 3.1 Perancangan Olah Dataset

Dataset ini memiliki detail tentang tindakan yang dilakukan penumpang saat menggunakan maskapai penerbangan. Menggunakan metode hubungan LRFM dan

algoritma *Clustering Expectation Maximization*, dataset memiliki berbagai atribut yang relevan untuk analisis segmentasi *Customer Lifetime Value*.

Dataset ini berjumlah 7988 entri data. Data ini terkait dengan satu transaksi atau keterlibatan konsumen dengan maskapai penerbangan. Kami akan mengkategorikan pelanggan menggunakan kumpulan data ini untuk lebih memahami kebiasaan dan preferensi pembelian pelanggan.

Berikut adalah atribut data yang relevan dengan penelitian ini:

1. MEMBER_NO: Kode unik untuk membedakan antara pelanggan satu dengan yang lainnya
2. LOAD_TIME: Menunjukkan tanggal saat data diambil atau dicatat
3. FFP_DATE: Menunjukkan tanggal pelanggan menjadi anggota dari maskapai penerbangan (*Frequent Flyer Program*)
4. LAST_TO_END: Merupakan jumlah hari terakhir sejak transaksi terakhir pelanggan hingga tanggal diambilnya data
5. FLIGHT_COUNT: Merupakan jumlah penerbangan yang telah dilakukan oleh pelanggan dalam jangka waktu tertentu
6. SEG_KM_SUM: Menyatakan total jarak penerbangan yang telah ditempuh oleh pelanggan (dalam satuan kilometer) dalam jangka waktu tertentu

3.2.1 Preprocessing data

Untuk memastikan bahwa data yang digunakan dalam analisis memiliki kualitas dan akurasi yang tinggi, persiapan data dilakukan untuk menghilangkan potensi masalah dari data, seperti nilai yang hilang, outlier data, atau duplikasi data.

Kami menggunakan metode *preprocessing* dalam penelitian ini (Alasadi & Bhaya, 2017). Langkah langkah berikut termasuk dalam teknik *preprocessing* ini:

1. Penanganan *Missing Values*

Beberapa teknik dapat digunakan untuk menangani nilai yang hilang. Mengisi angka yang hilang dengan statistik dari data yang ada merupakan salah satu cara yang sering digunakan. Sebagai gambaran, *missing value* dapat diganti dengan nilai rata-rata (mean) atau nilai median atribut tersebut.

$$Mean = \sum_{i=1}^{n} X_i \tag{3.1}$$

Yang mana:

Mean = nilai rata-rata dari atribut yang bersangkutan

X_i = nilai pada data ke-i

n = jumlah data yang tersedia untuk atribut tersebut

$$Median = \frac{X_{(n+1)/2} + X_{(n+1)/(2+1)}}{2} \tag{3.2}$$

Yang mana:

Median = nilai tengah dari atribut yang bersangkutan

$X_{(n+1)/2}$ dan $X_{(n+1)/(2+1)}$ = data ke-n dan data ke-n+1 setelah data urut

Mempertahankan integritas data dan meminimalkan bias dalam penelitian keduanya dapat dicapai dengan mengganti nilai yang hilang dari statistik dari data yang tersedia. Teknik pengisian *missing value* harus dipilih, namun sesuai dengan karateristik data dan tujuan analitis yang diinginkan.

2. Deteksi dan Penanganan *Outlier*

Untuk menemukan dan menangani nilai ekstrem yang sangat berbeda dari nilai lain dalam kumpulan data, deteksi dan penanganan outlier merupakan langkah penting dalam persiapan data. Nilai yang signifikan atau tidak biasa, yang dikenal sebagai *outlier*. Teknik ini dapat membuat analisis data menjadi miring. Penanganan outlier berusaha memperlakukan data tersebut agar tidak mengganggu hasil kajian utama. Deteksi outlier digunakan untuk mengidentifikasi data yang mencurigakan.

3. Deduplikasi data

Untuk menjamin integritas dan kualitas keluaran analitik, deduplikasi data sangat penting. teknik dapat memastikan bahwa analisis dilakukan pada kumpulan data yang bersih dan terstruktur dengan baik serta menghasilkan hasil yang lebih andal dan kredibel dengan menghapus data duplikat. Sebelum melakukan analisis tambahan, *preprocessing* data seringkali dimulai dengan langkah pertama yang krusial yang disebut deduplikasi data.

3.2.2 Normalisasi Dan Modeling Data

Tahap normalisasi data sangat penting untuk preprocessing analisis segmentasi *Customer Lifetime Value*. Agar atribut LRFM (*Length, Recency, Frequency, Monetary*) dapat diproses dan dibandingkan secara adil, normalisasi data dilakukan untuk memastikan memiliki skala yang konsisten dan setara. Jarak terbang, misalnya, dihitung dalam bilangan bulat sedangkan frekuensi transaksi dicatat dalam kilometer. Setiap atribut mungkin memiliki unit atau rentang nilai yang berbeda. Tanpa normalisasi, fitur skala yang lebih besar mungkin memiliki

dampak yang lebih kuat pada analisis dan segmentasi, mungkin menyebabkan hasil yang bias. Dengan normalisasi, data akan diubah ke dalam skala yang setara, seperti rentang nilai antara 0 hingga 1, sehingga memungkinkan setiap atribut untuk memberikan kontribusi yang seimbang dalam analisis.

Analisis perbandingan dan segmentasi menjadi lebih akurat dan mudah dipahami melalui normalisasi data. Hasil analisis juga lebih mudah dibaca ketika sudah dinormalisasi karena setiap atribut akan diberi bobot yang sama ketika klien dibagi menjadi segmen-segmen yang homogen. Algoritma *Clustering Expectation Maximization* dan metode hubungan LRFM dapat menghasilkan segmentasi pelanggan yang lebih akurat dengan dataset yang dinormalisasi, memungkinkan perusahaan untuk lebih memahami perilaku pelanggan, menyusun strategi pemasaran, dan mengelola pelanggan secara keseluruhan. Normalisasi data adalah prosedur prapemrosesan penting yang menjamin integritas data, meminimalkan bias, dan meningkatkan akurasi analitik untuk menghasilkan data yang berwawasan bagi bisnis.

1. Length

Dengan menghitung selisih antara tanggal penerbangan terakhir (LOAD_TIME) dan tanggal keanggotaan (FFP_DATE) untuk setiap pelanggan, normalisasi untuk properti *length* dalam data dapat diselesaikan. Angka *length* sekarang akan menjadi nilai relatif yang menunjukkan panjang rata-rata langganan setiap maskapai pelanggan sebagai akibat dari normalisasi ini.

Dataset setalah dilakukan preprosesing pada Tabel 3.1 menyajikan nilai yang dituliskan tersebut selanjutnya akan dilakukan perhitungan parameter *length*

dengan cara merubah tipe data *datetime* menjadi tipe data *Unix Timestamp* supaya bisa dilakukan operasi matematika.

Tabel 3.1 Dataset Parameter Length

MEMBER_NO	FFP_DATE	LOAD_TIME
27355	9/19/2011	3/31/2014
4032	9/18/2012	3/31/2014
12913	3/18/2009	3/31/2014
58465	9/6/2012	3/31/2014

Pemrosesan dan perbandingan nilai waktu di antara pelanggan akan mendapat manfaat dari data yang dinormalisasi, yang memfasilitasi identifikasi segmen klien berdasarkan fitur LRFM yang seragam. Selain itu, penerapan teknik analisis berbasis waktu yang lebih canggih dan tepat dimungkinkan dengan representasi waktu dalam bentuk stempel *Unix Timestamp*. Normalisasi data ke *Unix Timestamp* adalah prosedur prapemrosesan penting yang memastikan kualitas dan kebenaran data untuk mendapatkan informasi yang lebih mendalam tentang perilaku pelanggan.

Tabel 3.2 Hasil Konversi ke Unix Timestamp

MEMBER_NO	FFP_DATE	LOAD_TIME
27355	1316390400	1396224000
4032	1347926400	1396224000
12913	1237334400	1396224000
58465	1346889600	1396224000

Setelah menentukan parameter *length* dengan mengurangi nilai *Unix Timestamp* dari FFP_DATE dengan LOAD_TIME. Jumlah mewakili jumlah detik yang telah berlalu antara dua tanggal dalam representasi *Unix Timestamp*.

$$\text{Length} = \text{abs}(\text{FFP\_DATE} - LOAD\_TIME) \quad (3.4)$$

Parameter *length* dihitung menggunakan rumus *length* = abs(FFP_DATE - LOAD_TIME). Nilai *length* dalam rumus ini ditentukan dengan mengurangkan

nilai absolut representasi *Unix Timestamp* untuk FFP_DATE (tanggal keanggotaan) dan LOAD_TIME (tanggal penerbangan terakhir). Panjang langganan pelanggan ke perusahaan, terlepas dari apakah mereka bergabung sebelum atau setelah tanggal penerbangan terakhir, diwakili oleh nilai *length*, yang selalu positif saat fungsi absolut (*abs*) digunakan.

Tabel 3.3 Hasil Normalisasi atribut FFP_DATE dan LOAD_TIME

MEMBER_NO	Length
27355	79833600
4032	48297600
12913	158889600
58465	49334400

Setelah mendapatkan nilai *length* selanjutnya nilai tersebut harus dinormalisasi dengan teknik *Min-Max.* Untuk mengatasi skala angka yang besar dapat menerapkan metode *Min-Max Scaling*. Metode ini akan membantu mengubah rentang nilai hasil perhitungan menjadi nilai antara 0 hingga 1, sehingga memperkecil perbedaan skala dan memudahkan perbandingan antara atribut.

$$Scaled = \frac{length - min(length)}{max(length) - min(length)} \quad (3.5)$$

Yang mana :

min(length) = Nilai minimum *length* dalam seluruh data

max(length) = Nilai maksimum *length* dalam seluruh data

$$MinMax = Scaled * (new_{max} - new_{min}) + new_{min} \quad (3.6)$$

Yang mana:

Scaled = Hasil perhitungan dari rumus 3.5

new_{max} = Nilai maksimum dari atribut *length* dalam seluruh data

new_{min} = Nilai minimum dari atribut *length* dalam seluruh data

Untuk mengatasi nilai 0 pada hasil *Min-Max Scaling*, dapat menggunakan teknik yang disebut *Min-Max Scaling with Feature Scaling*. Teknik ini memungkinkan kita untuk menghindari nilai 0 pada hasil scaling dan menjaga konsistensi dalam data. Dengan menerapkan *Min-Max* Scaling, nilai *Length* akan diubah menjadi nilai yang lebih kecil dan setara antara 0 hingga 1. Hal ini akan memudahkan analisis dan perbandingan antara pelanggan.

Tabel 3.4 Hasil Normalisasi MinMax Parameter Length

MEMBER_NO	Length
27355	0.389
4032	0.110
12913	1
58465	0.119

2. Recency

Dengan menggunakan normalisasi *recency* dapat lebih memahami perilaku pelanggan berdasarkan lamanya waktu sejak transaksi terakhir mereka dan mengidentifikasi pelanggan yang lebih aktif dan berpotensi untuk memberikan kontribusi lebih kepada perusahaan.

Tabel 3.5 Dataset Parameter Recency

MEMBER_NO	LAST_TO_END
27355	549
4032	57
12913	385
58465	288

Selain itu, normalisasi *recency* membantu dalam pengembangan strategi pemasaran yang lebih efektif dengan menargetkan pelanggan dengan *recency* yang lebih rendah dengan penawaran promosi atau program loyalitas, dan mendekati pelanggan dengan *recency* yang lebih tinggi dengan pendekatan yang berbeda

untuk mengundang mereka kembali berinteraksi dan bertransaksi dengan perusahaan.

Tabel 3.6 Hasil Normalisasi MinMax Parameter Recency

MEMBER_NO	Recency
27355	1
4032	0.01
12913	0.81
58465	0.62

3. Frequency

Parameter *frequency* berguna dalam analisis segmentasi karena dapat membantu mengidentifikasi konsumen yang paling berkontribusi terhadap pendapatan perusahaan dan mengevaluasi seberapa terlibat dan loyal pelanggan terhadap merek atau layanan yang diberikan. Pelanggan dengan *frequency* tinggi cenderung setia kepada perusahaan dan menjadi pelanggan berulang yang menguntungkan.

Tabel 3.7 Dataset Parameter Frequency

MEMBER_NO	FLIGHT_COUNT
27355	2
4032	3
12913	3
58465	4

Nilai *frequency* juga harus dinormalisasi menggunakan teknik *Min-Max Scaling* agar dapat dibandingkan dengan atribut lain dalam analisis segmentasi. Nilai *frequency* akan diubah oleh normalisasi ini menjadi nilai relatif yang jatuh antara 0 dan 1. Berdasarkan *frequency* transaksi yang terlah mereka lakukan, normalisasi *frequency* ini akan membantu dalam perbandingan dan analisis karakteristik klien di masa mendatang.

Tabel 3.8 Hasil Normalisasi MinMax Parameter Frequency

MEMBER_NO	FLIGHT_COUNT
27355	0,01
4032	0,12
12913	0,12
58465	0,25

4. Monetary

Parameter *monetary* merupakan karakteristik yang menjelaskan nilai *monetary* keseluruhan dari semua transaksi atau pembelian yang dilakukan oleh setiap pelanggan. Faktor ini sangat penting dalam menentukan seberapa besar setiap konsumen berkontribusi secara finansial untuk bisnis. Pelanggan dengan nilai finansial yang cukup besar telah terlibat dalam transaksi dengan perusahaan yang memiliki kemungkinan kuat untuk meningkatkan nilai tersebut di masa mendatang.

Tabel 3.9 Dataset Parameter Monetary

MEMBER_NO	SEG_KM_SUM
27355	2005
4032	2301
12913	2397
58465	2852

Prameter *monetary* juga dapat dinormalisasi, seperti parameter lainnya, untuk menyamakan skala data selama analisis segmentasi. Dengan menggunakan teknik normalisasi seperti penskalaan *Min-Max Scaling*, nilai *monetary* dapat diubah menjadi nilai relatif antara 0 dan 1. Perbandingan dan analisis lebih lanjut atribut klien berdasarkan nilai moneter transaksi mereka akan menjadi lebih mudah berkat hasil normalisasi ini.

Tabel 3.10 Hasil Normalisasi MinMax Parameter Monetary

MEMBER_NO	SEG_KM_SUM
27355	0.01
4032	0.28
12913	0.40
58465	0.99

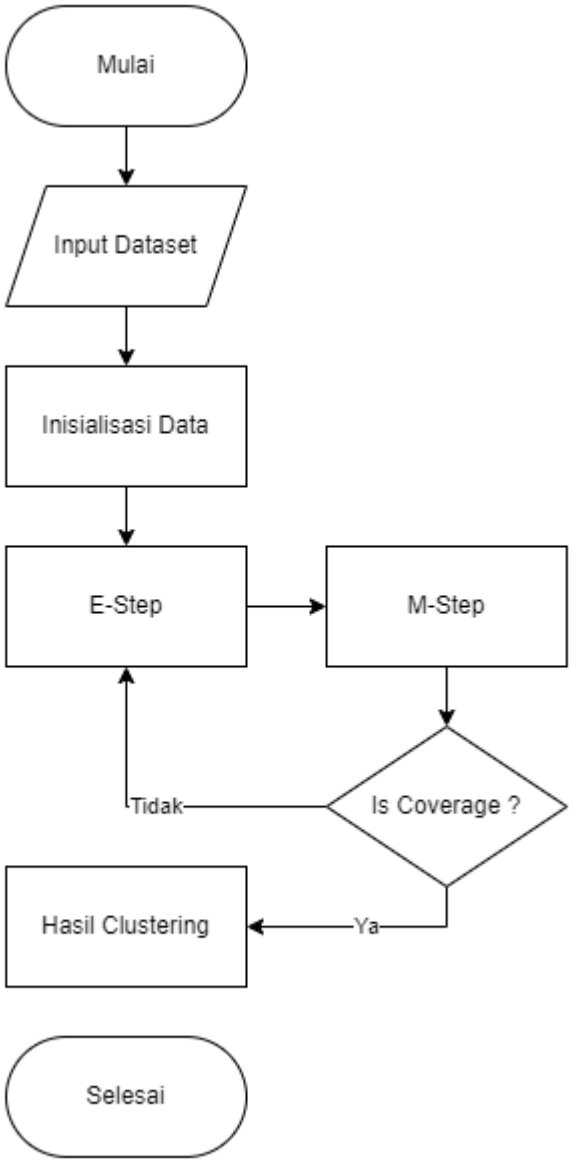
3.2.3 Clustering Expectation Maximization

Pada proses *clutering* menggunakan *expectation maximization* memiliki peran penting dalam analisis data dan pemodelan probabilistik. Proses *expectation maximization* adalah metode iteratif yang digunakan untuk memperkirakan parameter model statistik. Untuk mencapai konvergensi ke solusi ideal, pendekatan ini menggabungkan fase *expectation* (*E-step*) dan *maximization* (*M*-step).

Algoritma *expectation maximization* masih menggunakan parameter saat ini di *E-step* untuk menentukan nilai yang diharapkan dari distribusi variabel tersembunyi. Prosedur ini membantu memperbaharui pemahaman data yang tidak sepenuhnya dapat diamati. Langkah selanjutnya dalam *M-step* adalah memaksimalkan fungsi *log-likelihood* menggunakan parameter model yang telah direvisi. Parameter yang memiliki nilai *log-likelihood* tertinggi dipilih sebagai estimasi baru untuk parameter model pada langkah ini.

Algoritme *expectation maximization* berkembang menuju konvergensi dan menghasilkan estimasi parameter yang lebih tepat untuk model probabilistik dengan iterasi iterasi antara *e-step* dan *m-step*. Keuntungan mendasar dari metode *expectation maximization* adalah dapat mengelola data yang hilang atau variabel tersembunyi, yang menjadikannya alat yang sangat berharga dalam berbagai

aplikasi analisis data, seperti di bidang pemodelan statistik, pemrosesan bahasa alami, dan pengenalan pola. Proses *expectation maximization* melibatkan beberapa perhitungan seperti ditunjukkan pada Gambar 3.2 sebagai berikut.



Gambar 3.2 Diagram alur expectation maximization

Algoritma EM terdiri dari dua langkah untuk setiap iterasi. Langkah ekspektasi (*E-Step*) merupakan langkah pertama, sedangkan langkah maksimalisasi

(*M-Step*) merupakan langkah kedua. Proses E-Step digunakan untuk mencari suatu fungsi yaitu ekspetasi dari fungsi likehood yang dinotasikan dengan berikut.

$$eval(d) = \frac{\pi \,.\, pdf(d, \mu_2, \sum_2)}{\pi \,.\, pdf(d, \mu_2, \sum_2) + (1 - \pi) \,.\, pdf(d, \mu_1, \sum_1)} \quad (3.6)$$

Yang mana

eval(d) = Nilai ekspektasi untuk data

π = Bobot dari komponen kedua dalam campuran Gaussian.

pdf(d, μ, $\sum$) = Fungsi densitas probabilitas multivariat Gaussian untuk data *(d)* dengan rata-rata *(π)* dan matriks kovariansi ($\sum$)

$\mu_1, \sum_1$ = Rata-rata dan matriks kovariansi untuk komponen pertama dalam campuran Gaussian

$\mu_2, \sum_2$ = Rata-rata dan matriks kovariansi untuk komponen kedua dalam campuran Gaussian

Dengan menggunakan ekspektasi yang diberikan pada proses *E-Step* sebagai dasar, pada proses *M-Step* akan memaksimalkan fungsi *likelihood*. Persamaan *likelihood* setiap parameter untuk prosedur *E-Step* untuk dihasilkan. Fungsi m-step dinotasikan dengan berikut.

Perbaharui parameter rata-rata ($\mu_1 \; dan \; \mu_2$)

$$\mu_1 = \frac{\sum_{i=1}^{N}(1 - eval[i]) \,.\, d[i]}{\sum_{i=1}^{N}(1 - eval[i])} \quad (3.7)$$

$$\mu_2 = \frac{\sum_{i=1}^{N} eval[i] \,.\, d[i]}{\sum_{i=1}^{N} eval[i]} \quad (3.8)$$

Yang mana

$\mu_1 \, dan \, \mu_1$ = Parameter rata-rata untuk komponen pertama dan kedua dalam campuran Gaussian.

N = Jumlah data dalam set Anda.

eval[i] = Nilai ekspektasi yang telah dihitung dalam langkah Ekspektasi untuk data ke-i

D[i] = Data ke-i dalam set data Anda.

Perbaharui parameter matriks kovariansi ($\sum_1 \, dan \, \sum_2$)

$$\sum_1 = \frac{\sum_{i=1}^{N} (1 - eval[i]) \,.\, (d[i] - \mu_1) \,.\, (d[i] - \mu_1)^T}{\sum_{i=1}^{N} (1 - eval[i])} \quad (3.9)$$

$$\sum_2 = \frac{\sum_{i=1}^{N} (eval[i]) \,.\, (d[i] - \mu_1) \,.\, (d[i] - \mu_1)^T}{\sum_{i=1}^{N} (eval[i])} \quad (3.10)$$

Yang mana

$\sum_1 \, dan \, \sum_1$ = Matriks kovariansi untuk komponen pertama dan kedua dalam campuran Gaussian.

$\mu_1 \, dan \, \mu_1$ = Rata-rata yang telah diestimasi sebelumnya.

eval[i] = Nilai ekspektasi yang telah dihitung dalam langkah Ekspektasi untuk data ke-i.

d[i] = Data ke-i dalam set data Anda.

$(d[i] - \mu_1)^T$ = Vektor yang merupakan selisih antara data d[i] dan rata-rata yang sesuai.

Perbarui Parameter Bobot (π)

$$\pi = \frac{\sum_{i=1}^{N}(eval[i])}{N} \quad (3.11)$$

Yang mana

π = Bobot (weight) dari komponen kedua dalam campuran Gaussian.

N = Jumlah data dalam set Anda.

eval[i] = Nilai ekspektasi yang telah dihitung dalam langkah Ekspektasi untuk data ke-i.

Langkah *E-Step* dan *M-Step* diulang hingga mencapai estimasi parameter yang konvergen atau tidak mengalamai banyak perubahan.

3.3 Skenario Pengujian

3.3.1 Pengujian Software

Sebelum melakukan pengujian hasil *cluster*, langkah awal adalah menginstal perangkat lunak atau alat yang diperlukan untuk proses clustering, memastikan bahwa perangkat lunak berfungsi dengan baik dan dapat diakses, melakukan pengujian fungsionalitas perangkat lunak, serta memverifikasi kemampuan perangkat lunak untuk mengolah data dan menghasilkan hasil *cluster*.

3.3.2 Pengujian Parameter

Setelah memastikan perangkat lunak berfungsi dengan baik, langkah berikutnya adalah menguji parameter data yang digunakan dalam proses *clustering*. Dengan mengidentifikasi parameter seperti jumlah *cluster* yang optimal, jenis matrik jarak yang paling cocok, dan metode inisialisasi *cluster* yang efektif. Pengujian ini melibatkan eksperimen dengan berbagai kombinasi parameter untuk menemukan pengaturan terbaik.

3.3.3 Pengujian Hasil

Setelah persiapan data dan perangkat lunak, langkah selanjutnya adalah menjalankan proses *clustering* dengan perangkat lunak yang telah diinstal, mengevaluasi hasil *clustering* menggunakan metrik yang sesuai seperti Silhouette Score, menganalisis hasil *cluster* untuk mengidentifikasi pola yang muncul, membandingkan hasil *clustering* dengan solusi referensi jika ada, dan menguji kestabilan hasil *clustering* melalui beberapa percobaan.

Jumlah *cluster* yang ideal dapat ditentukan berdasarkan beberapa pertimbangan, seperti karakteristik data, tujuan penelitian, dan kesesuaian dengan hasil penelitian lain. Pengujian hasil *cluster* dapat dilakukan dengan menggunakan berbagai metode, seperti *Silhouette Score*.

BAB IV

UJI COBA DAN PEMBAHASAN

Implementasi sistem merupakan tahap pembuatan aplikasi ke dalam bentuk perangkat lunak sesuai dengan hasil analisis yang telah dilakukan untuk mengidentifikasi kekurangan maupun kelebihan pada aplikasi untuk selanjutnya diadakan perbaikan aplikasi atau sistem.

Tujuan implementasi sistem adalah menyederhanakan desain telah diselesaikan pada sistem, guna memungkinkan orang untuk berkontribusi penyempurnaan sistem dilakukan agar sistem menjadi lebih baik.

4.1 Implementasi

Implementasi merupakan suatu prosedur yang dapat mengubah atau penyususnan strategi menjadi sebuah variabel dalam mencapai tujuan atau sasaran tertentu. Implementasi juga dapat diartikan sebagai proses mengubah rancangan ke dalam variabel pemrograman. Pada implementasi ini akan mencangkup lingkup perangkat keras, lingkup perangkat lunak dan implementasi program klaster data pengguna maskapai penerbangan.

4.1.1 Ruang Lingkup Perangkat Keras

Perangkat keras (*hardware*) yang digunakan dalam program klasterisasi data pengguna maskapai penerbangan, yaitu Laptop MSI dengan prosesor Intel Core i7-8750H @ 2.20 Ghz, Ram 16 GB.

4.1.2 Ruang Lingkup Perangkat Lunak

Perangkat lunak (*software*) yang digunakan dalam program klasterisasi data pengguna maskapai penerbangan, sebagai berikut

1. Sistem Operasi Windows 11
2. IDE PHPStorm
3. PHP 8.3
4. IDE PyCharm
5. Python 3.11.9
6. Google chrome
7. MariaDb

4.1.3 Implementasi Program Clustering

Pada implementasi program klastering ini merujuk pada bab 3, dalam bab tersebut sudah dijelaskan mengenai proses atau tahapan yang dilakukan dalam program klasterisasi yang diawali dengan menginputkan atau memasukkan data dari dataset, menentukan parameter yang diperlukan, *preprosesing* data, *cleaning* data. Tahapan tahapan tersebut akan dibahas lebih jelas pada bahasan selanjutnya tentang implementasi dari tahapan-tahapan tersebut. Pada implementasi program *clustering* dengan algoritma *expectation maximization clustering* dengan model *LRFM* menggunakan *framework* laravel sebagai render antar muka dan *framework* flask sebagai *engine clustering*.

a. Implementasi Dataset

Tahapan pertama dalam implementasi adalah mengumpulkan semua data pada program. *Dataset* yang akan digunakan dibentuk dari 5 kolom yang ada di

dataset pada *Kaggle* yang kemudian diimpor dalam bentuk csv yang nantinya akan di simpan pada *database*. Untuk menghubungkan dataset ke database dapat menggunakan *package* matweb laravel untuk pengolahan data seperti membaca data dalam format csv ataupun sesuai dengan yang akan digunakan. Setelah mengimpor matweb dan *package* lainnya yang akan dibutuhkan dalam menjalakan program ini. Kemudian untuk memuat dataset dari csv ke dalam *database*, menggunakan kode php sebagai berikut.

```
public function import($importRequest): bool
{
      if ($importRequest->hasFile('file')) {
            $importRequest->file('file')->storeAs('public',
'flight.csv');
            dispatch(new
FlightImportJob(storage_path('app/public/flight.csv')));
            Konfigurasi::query()
                  ->where('id', '=',
Konfigurasi::IS_DONE_IMPORT)
                  ->update(['value' => 1]);
            return true;
      }
}
```

Pada kode tersebut menggunakan fungsi 'dispatch()' untuk menjalankan antrian membaca seluruh data yang ada di *dataset* dan path yang terdapat pada fungsi tersebut sesuai dengan *directory dataset* berada, karena dataset diletakkan dalam *drive* lokal yang bertujuan agar tidak salah dalam memilih *dataset* yang akan digunakan, sehingga *dataset* yang ada di *drive* akan tetap ditempat yang sama dan tidak akan hilang.

Kemudian menghubungkan olah dataset ke dalam databse agar nantinya data tersebut gampang dicari dan di olah dengan tahapan berikutnya.

```
public function collection(Collection $collection): void
{
      $datas = [];
      foreach ($collection as $row) {
            $ifExist = Flight::where('member_no',
$row['member_no'])->first();
            if ($ifExist) {
                  continue;
            }

            ....

            $datas[] = [
                  'member_no'         => $row['member_no'],
                  'fpp_date'          => $ffp_date,
                  'first_flight_date' => $first_flight_date,
                  ....
            ];
      }

      Flight::insert($datas);
}
```

Pada kode berikut data yang ada di *dataset* akan divalidasi jika data tersebut ada pada *database* maka baris tersebut akan diabaikan dan akan dilanjut dengan baris berikutnya. Kemudian data yang sudah diolah akan di simpan pada *database*.

b. Implementasi Preprosesing Data

Dalam implementasi *preprosesing* data terdapat 3 tahapan, yaitu deteksi duplikasi data, deteksi *outlier* dan deteksi data yang tidak lengkap.

```
public function duplicateDetection(): bool
{
      if (Konfigurasi::find(Konfigurasi::CHECK_DUPLICATE)-
>value) {
            return false;
      }

      $duplicate = $this->model::query()
            ->whereRaw('flight_fix_id not in (select
min(flight_fix_id) from flight_fix group by member_no)')
            ->get();

      if ($duplicate->count() > 0) {
            $duplicate->each(function($item) {
                  $item->delete();
            });
      }

      Konfigurasi::query()
            ->where('id', '=',Konfigurasi::CHECK_DUPLICATE)
            ->update(['value' => 1]);

      return true;
}
```

Pada potongan kode tersebut, aplikasi membaca konfigurasi dahulu sebelum eksekusi cek duplikasi data, jika pada konfigurasi sudah pernah melakukan cek duplikasi data maka tahap ini akan dilewati. Pada prosesnya program ini membuat sebuah *query* pada *database* yang mengambil semua data jika memiliki kesamaan lebih dari satu. Jika ditemukan sebuah data yang memiliki kesamaan lebih dari satu maka akan dihapus dan disisakan satu.

Saat proses cek duplikasi data sudah berhasil dilakukan, maka database untuk konfirugasi cek duplikasi akan diperbaharui untuk melewati proses cek duplikasi untuk ke-dua kalinya. Kemudian kembalikan nilai *boolean* kepada antar muka untuk menampilkan pesan sukses.

```
public function handlingPreprosesing(Collection|FlightFix
$flights): bool
{
      return $this->outlierDetection($flights);
}
```

Pada proses deteksi *outlier* dilakukan ketika terdapat data dengan nilai yang terlalu tinggi atau terlalu rendah bagi kebanyakan data yang akan digunakan dalam klastering.

Deteksi *outlier* merupakan langkah penting dalam analisis data yang bertujuan untuk mengidentifikasi data yang memiliki nilai terlalu ekstrim, baik terlalu tinggi atau terlalu rendah, jika dibandingkan dengan mayoritas data yang ada. Proses ini sangat penting ketika data akan digunakan dalam metode *clustering*. Dalam konteks *clustering*, keberadaan *outlier* dapat mengganggu proses pengelompokan, sehingga hasil *cluster* yang dihasilkan menjadi tidak akurat atau tidak representatif.

```
protected function missingDetection(Collection|FlightFix
$flights): bool
{
      $data = collect();

      foreach ($flights as $flight) {
```

```
            if (!$flight->member_no || !$flight->fpp_date ||
!$flight->load_time || !$flight->last_to_end || !$flight-
>flight_count || !$flight->seg_km_sum) {
                  continue;
            }

            $data->push([
                  'member_no'    => $flight->member_no,
                  'fpp_date'     => $flight->fpp_date,
                  'load_time'    => $flight->load_time,
                  'last_to_end'  => $flight->last_to_end,
                  'flight_count' => $flight->flight_count,
                  'seg_km_sum'   => $flight->seg_km_sum,
            ]);
      }

      return $this->bulkInsert($data);
}
```

Pada fungsi ‘missingDetection()’ tersebut memiliki parameter ‘$flights’ yang mana parameter tersebut merupakan hasil dari alur deteksi *outlier* yang kemudian akan dilakukan cek data yang kosong. Dalam kode tersebut seluruh data akan diulang sebanyak jumlah baris dan akan dicek, ketika ada salah satu kolom yang tidak berisi maka data tersebut tidak akan disimpan pada *databse* yang akan dilakukan tahap selanjutnya.

c. Implementasi Normalisasi Dan Modeling Data

Pada implementasi ini yaitu memproses data yang memiliki tipe data yang berbeda menjadi seragam. Dalam proses data terdapat dua tahap, yaitu normalisasi data dan *scaling min max*.

```
public function jobNormalisasi(Collection $flightFix): void
{
      $result = Collection::make();
      foreach ($flightFix as $data) {
            $fppDate  = $data->fpp_date;
            $loadTime = $data->load_time;

            $length    = abs(strtotime($fppDate) -
strtotime($loadTime));
            $recency   = $data->last_to_end;
            $frequency = $data->flight_count;
            $monetary  = $data->seg_km_sum;

            $result->push([
                  'member_no' => $data->member_no,
                  'length'    => $length,
                  'recency'   => $recency,
                  'frequency' => $frequency,
                  'monetary'  => $monetary,
            ]);
      }

      NormalisasiTemp::query()->insert($result->toArray());
}
```

Pada fungsi ‘jobNormalisasi()’ mengubah data pelanggan maskapai kedalam model LRFM (*Length, Recency, Frequency, Monetary*). Setiap baris dalam database akan diulang sebanyak data dan disimpan ke dalam memori untuk pengolahan tiap-tiap jenis model. Label member_no diambil dari variabel $data->member_no. Model *length* dari variabel $fppDate dikurangi $loadTime yang sebelumnya sudah diubah dari tipe data tanggal kedalam tipe data bilangan bulat.

Model *recency* diambil dari variabel $data->last_to_end. Model *frequency* diambil dari variabel $data->flight_count. Sedangkan untuk model *monetary* diambil dari variabel $data->seg_km_sum.

```
public function handle(): void
    {
        $normalisasiTemp = NormalisasiTemp::all();

        $min = [];
        $max = [];

        [$min, $max] = $this->evaluatedMinMax($normalisasiTemp,
$min, $max);

        $newNormalisasiTemp = Collection::make();

        $normalisasiTemp->each(function(NormalisasiTemp $item)
use ($min, $max, $newNormalisasiTemp) {
            $length    = ($item->length - $min['length']) /
($max['length'] - $min['length']);
            $recency   = ($item->recency - $min['recency']) /
($max['recency'] - $min['recency']);
            $frequency = ($item->frequency - $min['frequency'])
/ ($max['frequency'] - $min['frequency']);
            $monetary  = ($item->monetary - $min['monetary']) /
($max['monetary'] - $min['monetary']);

            $newNormalisasiTemp->push([
                'member_no' => $item->member_no,
                'length'    => $length,
                'recency'   => $recency,
                'frequency' => $frequency,
```

```
                'monetary'  => $monetary,
            ]);
        });

        [$min, $max] = $this-
>evaluatedMinMax($newNormalisasiTemp, $min, $max);

        $normalisasiFix = Collection::make();

        $newNormalisasiTemp->each(function(array $item) use
($min, $max, $normalisasiFix) {
            $length    = ($item['length'] * ($max['length'] -
$min['length'])) + $min['length'];
            $recency   = ($item['recency'] * ($max['recency'] -
$min['recency'])) + $min['recency'];
            $frequency = ($item['frequency'] *
($max['frequency'] - $min['frequency'])) + $min['frequency'];
            $monetary  = ($item['monetary'] * ($max['monetary']
- $min['monetary'])) + $min['monetary'];

            $normalisasiFix->push([
                'member_no' => $item['member_no'],
                'length'    => $length <= 0 ? 0.0111111 :
$length,
                'recency'   => $recency <= 0 ? 0.0111111 :
$recency,
                'frequency' => $frequency <= 0 ? 0.0111111 :
$frequency,
                'monetary'  => $monetary <= 0 ? 0.0111111 :
$monetary,
            ]);
        });
```

```
        $normalisasiFix->chunk(1000)->each(fn($data) =>
dispatch(new SaveNormalisasiDataJob($data->toArray())));
    }
```

Pada fungsi *scaling min max* ini, bertujuan untuk melakukan normalisasi data pada data sesudah diseragamkan seluruh data yang ada. Pertama, seluruh data dari *database* diambil dan disimpan dalam variabel $normalisasiTemp. Selanjutnya, fungsi evaluatedMinMax digunakan untuk menghitung nilai minimum dan maksimum dari atribut-atribut yang relevan (seperti *length*, *recency*, *frequency*, dan *monetary*). Setelah mendapatkan nilai minimum dan maksimum, data normalisasi baru dibuat dalam variabel $newNormalisasiTemp dengan melakukan transformasi pada setiap atribut berdasarkan nilai minimum dan maksimum yang telah dihitung sebelumnya. Setelah normalisasi awal, nilai minimum dan maksimum dievaluasi kembali untuk memastikan konsistensi. Data yang telah dinormalisasi disimpan dalam variabel $normalisasiFix dengan menambahkan nilai yang sudah dikembalikan ke skala aslinya, memastikan bahwa nilai yang sangat kecil tidak bernilai nol dengan menetapkan batas bawah 0.0111111. Akhirnya, data yang telah dinormalisasi dipecah menjadi bagian-bagian kecil (*chunk*) berukuran 1000 dan setiap bagian diproses melalui pekerjaan asinkron SaveNormalisasiDataJob untuk disimpan ke dalam basis data. Implementasi ini memastikan bahwa proses normalisasi dilakukan secara efisien dan hasilnya disimpan dengan cara yang terdistribusi.

d. Implementasi Klasterisasi Data (Expectation Maximization)

Pada implementasi terakhir ini merupakan inti dari program yang dibuat yaitu clustering. Pada proses *expectation maximization* disini melibatkan banyak

parameter yang sudah diperolah dari tahap sebelumnya dan juga nilai konfigurasi dari algoritma klastering. Hal pertama yang dilakukan dalam proses *expectation maximization* yaitu menentukan atau menghitung parameter yang optimal.

Dalam proses melibat dua *framework* dan bahasa pemrograman yang berbeda, untuk itu diperlukan sebuah *API* (*Application Programming Interface*) yaitu sebuah teknik penggabungan sebuah bahasa pemrograman yang berbeda agar dapat berjalan secara paralel.

```
public function clustering()
{
	$remote = app(Clustering::class);
	app(SilhoutteService::class)->calculateSilhoutte();

	$remote->cluster();
}

protected function post(string $url, ?array $data = []): array
{
	$response = $this->makeRequest()
		->timeout(3600)
		->withHeaders($this->headers)
		->post($url, $data);

	if ($response->failed()) {
		throw new Exception($response-
>json()['responseMessage'] ?? 'Remote service error');
	}

	return $response->json();
}
```

```
public function cluster(): void
{
      $this->post('/cluster');
}
```

Dalam potongan kode berikut fungsi ‘cluster()’ diinisialisasi dengan *service container* kelas *Clustering::class.* Pada kelas *Clustering::class* digunakan untuk memangging *endpoint* dari *framework flask* dengan konfigurasi sebagai *request* dengan timeout 1 menit dan menggunakan sebuah *header json*.

```
@app.route('/cluster', methods=['POST'])
def cluster():
    n_clusters = 4
    limit = 10000

    # Mengambil data dari database
    cluster_results =
Normalisasi.query.with_entities(Normalisasi.member_no,
Normalisasi.length,

Normalisasi.recency, Normalisasi.frequency,

Normalisasi.monetary).limit(limit).all()

    # Mengambil label
    labels = [result.member_no for result in cluster_results]

    # Mengonversi data dari string ke array numpy
    data_np = np.array(cluster_results)[:, 1:].astype(float)

    # Menerapkan algoritma EM
    em = EM_Clustering(n_clusters=n_clusters)
    em.fit(data_np)
```

```
    # Memprediksi cluster
    predicted_clusters = em.predict(data_np)

    # Menggunakan fungsi optimized_upsert
    optimized_upsert(labels, data_np, predicted_clusters)

    # Mengembalikan hasil clustering dalam format JSON
    return jsonify({'predicted_clusters': 'berhasil'})
```

Pada program *python* tersebut akan mengkalkulasi sebuah cluster menggunakan data dari *database* yang data tersebut sudah diolah dari tahapan sebelumnya. Dengan menggunakan konfigurasi 4 cluster dengan limit data 10000 baris dan *max_iter* sebanyak 500 yang artinya perulangan dalam perhitungan cluster akan berhenti jika sudah 500 atau sudah mencapai nilai optimal dan toleransi sebesar 0,001 yang artinya nilai error atau berhenti yang diharapkan yaitu 0,001.

Data hasil clustering akan disimpan pada tabel baru yaitu *ClusterResult* agar data yang sudah menjadi cluster tidak bersatu dengan data yang belum diolah. Setelah melakukan perhitungan pada e-step dan m-step pada algoritma *expectation maximization*, kemudian menentukan kriteria pemberhentian proses clustering. Untuk menentukan kriteria pemberhentian yaitu apabila nilai yang diharapkan sesuai atau lebih kecil dari nilai tolerance yang sudah ditentukan pada implementasi paramete maka iterasi akan diberhentikan. Begitu juga, apabila iterasi atau nilai error yang diharapkan sudah sesuai makai iterasi diberhentikan, namun jika tidak sesuai maka iterasi dilanjutkan sampat batas maksimal iterasi yang sudah ditentukan

4.2 Pembahasan

Pembahasan bertujuan untuk menyajikan atau menjelaskan hasil dari penelitian yang sudah dilakukan dari implementasi yang sudah dibuat sebelumnya. Berikut beberapa poin yang terdapat dalam pembahasan ini.

4.2.1 Hasil Cluster Terbaik Menggunakan Expectation Maximization

Pada proses klastering menggunakan algoritma *expectation maximization* yang sudah dilakukan sebelumnya, dapat menghasilkan *cluster-cluster* yang diinginkan. Pada menggunakan perhitungan *silhouette* atau matriks evaluasi yang digunakan untuk analisis *cluster* dan untuk menentukan jumlah *cluster* yang optimal dalam data mining.

```
public function calculateSilhoutte(): void
    {
        $data =
file_get_contents(database_path(silhoutte_score'));
        $json = json_decode($data, true);

        $table   = SilhoutteScore::query()->getModel()-
>getTable();
        $columns = Schema::getColumnListing($table);

        foreach (array_chunk($json ?: [], 1000) as $data) {
            $value = Arr::map($data, fn($item) =>
Arr::only($item, $columns));
            DB::table($table)
                ->insertOrIgnore($value);
        }
    }
```

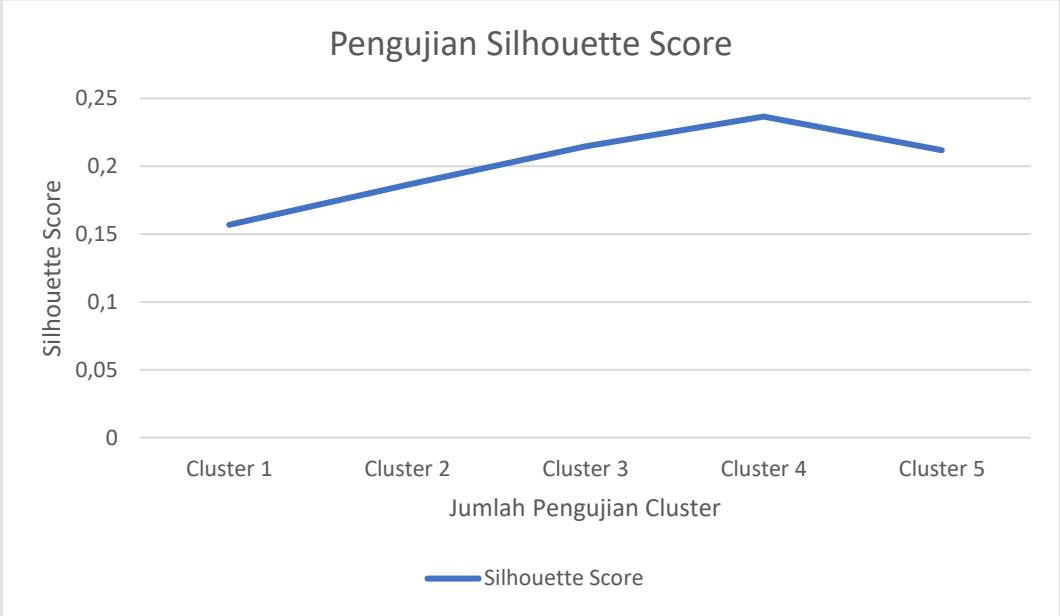
Pada kode diatas merupakan proses perhitungan nilai *silhouette* dan akan dimasukkan kedalam database. Proses tersebut akan otomatis berjalan ketika proses *clustering* akan dijalankan sehingga proses *clustering* akan pralel dengan perhitungan skor *silhouette*.

Tabel 4.1 Skenario Pengujian Silhouette Score

No	Data	N Cluster
1	Model LRFM	1
2		2
3		3
4		4
5		5

Pada tabel 4.1 skenario pengujian perhitungan silhouette score, pengujian dilakukan sebanyak 5 kali dengan rincian sebagai berikut:

1. Pengujian pertama dilakukan dengan melakukan variasi jumlah cluster sebesar 1 cluster dan mendapatkan nilai 0,15684
2. Pengujian kedua dilakukan dengan melakukan variasi jumlah cluster sebesar 2 cluster dan mendapatkan nilai 0,18629
3. Pengujian ketiga dilakukan dengan melakukan variasi jumlah cluster sebesar 3 cluster dan mendapatkan nilai 0,21459
4. Pengujian keempat dilakukan dengan melakukan variasi jumlah cluster sebesar 4 cluster dan mendapatkan nilai 0,23662
5. Pengujian kelima dilakukan dengan melakukan variasi jumlah cluster sebesar 5 cluster dan mendapatkan nilai 0,21187



Gambar 4.1 Hasil Perhitungan Score silhouette

Pada gambar 4.1, menunjukkan bahwa dengan membagi data menjadi 4 *cluster* merupakan paling optimal daripada yang lain. Dapat dilihat dalam grafik di atasmenunjukkan *elbow method* terjadi pada jumlah 4 cluster. Dengan total 4 *cluster* dapat memaksimalkan homogenitas internal setiap *cluster* dan meminimalkan heterogenitas antar *cluster*. Dengan uraian sebagai berikut, satu *cluster* dengan *score* 0,15684, dua *cluster* dengan *score* 0,18629, tiga *cluster* dengan *score* 0,21459, empat *cluster* dengan *score* 0,23662 sedangkan lima *cluster* dengan *score* 0,21187.

4.2.2 Pengujian Variasi Principal Component Analysis

Parameter principal component analysis mengacu pada jumlah kolom yang akan disederhanakan menjadi n kolom. Nilai parameter n yang akan di gunakan adalah nilai bilangan bulat mulai dari 14, 15, 16, 17, 18 dan 19. Pengujian dilakukan dengan melakukan iterasi variasi jumlah n parameter. Pengujian ini diharapkan

dapat menemukan nilai pca terbaik berdasarkan nilai *explained variance ratio* atau jumlah komponen optimal yang bisa mewakilkan semua parameter asli.

Tabel 4.2 Skenario pengujian Principal Component Analysis

No	Data Customer	Variasi PCA
1	Data Customer Flight	14
2		15
3		16
4		17
5		18
6		19

1. Pengujian pertama dengan 14 parameter variasi principal component analysis dari 20 parameter data awal.
2. Pengujian pertama dengan 15 parameter variasi principal component analysis dari 20 parameter data awal.
3. Pengujian pertama dengan 16 parameter variasi principal component analysis dari 20 parameter data awal.
4. Pengujian pertama dengan 17 parameter variasi principal component analysis dari 20 parameter data awal.
5. Pengujian pertama dengan 18 parameter variasi principal component analysis dari 20 parameter data awal.
6. Pengujian pertama dengan 19 parameter variasi principal component analysis dari 20 parameter data awal.

Tabel 4.3 Hasil Pengujian Variasi Principal Component Analysis

PCA	Explained Variance Ratio				
	Komponen 1	Komponen 2	Komponen 3	Komponen 4	Total
14	10.25%	22.15%	15.35%	20.15%	67.90%
15	22.10%	25.00%	18.35%	24.00%	89.45%
16	22.03%	20.13%	22.50%	19.87%	84.53%
17	20.50%	21.45%	18.75%	20.62%	81.32%
18	21.50%	22.00%	19.75%	22.40%	85.65%
19	21.25%	19.75%	23.24%	23.20%	87.44%

Dari pengujian variasi *principal component analysis* pada tabel 4.3 dapat dilihat bahwa dengan 15 variasi *principal component analysis* sudah dapat mewakili 20 parameter yang ada dengan total skor 89.45%. Dengan detil 22.10% pada komponen 1, 25.00% pada komponen 2, 18.35 pada komponen 3 serta 24.00% pada komponen 4.

Tabel 4.4 Data hasil cluster LRFM dengan Expectation Maximization

Member No	Parameter				K Cluster
	L	R	F	M	
11163	0,26302	0,26575	0,01111	0,32597	0
30765	0,14518	0,86027	0,12500	0,28413	3
10380	0,86914	0,50822	0,01111	0,48509	3
16372	0,24837	0,55753	0,25000	0,32845	0
22761	0,62565	0,23151	0,12500	0,25431	0
11163	0,21191	0,11507	0,25000	0,67965	2
34330	0,00260	0,39178	0,01111	0,34715	2
1761	0,48145	0,90411	0,12500	0,46012	3
…	…	…	…	…	…
16415	0,64486	0,00137	0,01111	0,50431	1

Hasil dari percobaan yang dilakukan untuk mendapatkan hasil cluster dengan dataset yang sama menggunakan *platform* Kaggle dapat dilihat pada tabel 4.1. Pada perhitungan menggunakan kaggle dengan 7988 didapatkan 4 buah *cluster* dengan masing masing 655 data pada *cluster* 1, 2866 data pada *cluster* 2, 980 data pada *cluster* 3 dan 3487 pada *cluster* 0. Dengan label setiap cluster sebagai berikut,

cluster 0 sebagai *Potential*, *cluster* 1 sebagai *Important*, *cluster* 2 sebagai *general & Low Value* sedangkan *cluster* 3 sebagai *Loyal*.

4.2.3 Pembahasan Hasil Penelitian

Pada perhitungan dan hasil yang sudah didapatkan, menjelaskan bahwa program ini dapat berjalan dengan baik. Total *dataset* 7988 dapat dieksekusi dalam waktu 18,3 detik. Pada tabel hasil *cluster* terlihat ada 4 jenis *cluster* yang dibentuk berdasarkan *dataset* yang sudah ditentukan sebelumnya dan sudah melalui tahap *preprosesing* dan normalisasi data saat akan menjalankan proses klastering dengan algoritma *expectation maximization*. Dengan hasil label *cluster* 0 sebagai *loyal*, *cluster* 1 sebagai *potential*, *cluster* 2 sebagai *low value* sedangkan *cluster* 3 sebagai *important*.

Dari seluruh data yang diklasterisasi menghasilkan 1135 data pelanggan pada cluster 0 yang artinya 1135 pelanggan maskapai memiliki nilai *loyal*, sedangkan cluster 1 terdapat 2873 data pelanggan yang artinya terdapat 2873 pelanggan maskapai yang memiliki nilai *potential*, untuk cluster 2 terdapat 3149 data pelanggan yang memiliki nilai *low value*, dan terdapat 731 data pelanggan pada cluster 4 yang artinya 731 pelanggan tersebut memiliki nilai *important*.

Dari hasil *cluster* dapat dilihat algoritma *expectation maximization* mengelompokkan dengan pembobotan yang akan dilanjutkan dengan derajat keanggotaan pada algoritma tersebut. Untuk membandingkan data hasil *cluster* dengan algoritma *expectation maximization* dengan variasi *principal component analysis* dapat dilihat dari tabel perbandingan berikut.

Tabel 4.5 Perbandingan hasil clustering EM dan PCA Variasi 15

No Member	Parameter				EM	PCA 15
	L	R	F	M		
11163	0,26302	0,26575	0,01111	0,32597	General & Low Value	Potential
30765	0,14518	0,86027	0,12500	0,28413	Important	Potential
10380	0,86914	0,50822	0,01111	0,48509	Loyal	Important
16372	0,24837	0,55753	0,25000	0,32845	Potential	Potential
22761	0,62565	0,23151	0,12500	0,25431	Important	Loyal
35869	0,21191	0,11507	0,25000	0,67965	General & Low Value	General & Low Value
12054	0,00260	0,39178	0,01111	0,34715	General & Low Value	General & Low Value
1761	0,48145	0,90411	0,12500	0,46012	General & Low Value	Loyal
…	…	…	…	...	…	
16415	0,64486	0,00137	0,01111	0,50431	Important	Loyal

4.3 Integrasi Islam

Proses pengelompokan melibatkan pemisahan kumpulan data menjadi banyak kelompok. data yang dikelompokkan bersama dengan data lain yang memiliki sifat yang sama. Dalam Surat Al-Fathir ayat 32 Al-Qur'an menjelaskan bagaimana mengelompokkan orang berdasarkan sikapnya. Ayat 32 Surat Al-Fathir adalah sebagai berikut:

ثُمَّ اَوْرَثْنَا الْكِتٰبَ الَّذِيْنَ اصْطَفَيْنَا مِنْ عِبَادِنَاۚ فَمِنْهُمْ ظَالِمٌ لِّنَفْسِهٖۚ وَمِنْهُمْ مُّقْتَصِدٌۚ وَمِنْهُمْ سَابِقٌۢ بِالْخَيْرٰتِ بِاِذْنِ اللّٰهِۗ ذٰلِكَ هُوَ الْفَضْلُ الْكَبِيْرُۗ

Yang artinya: "*Kemudian Kitab itu Kami wariskan kepada orang-orang yang Kami pilih di antara hamba-hamba Kami, lalu di antara mereka ada yang menganiaya diri mereka sendiri dan di antara mereka ada yang pertengahan dan diantara mereka ada (pula) yang lebih dahulu berbuat kebaikan dengan izin Allah. Yang demikian itu adalah karunia yang amat besar.*" (Surat Al-Fathir ayat 32)

Dalam Surat An-Nisa ayat 29 menegaskan perlunya menjaga prinsip-prinsip moral dalam perdagangan serta perlunya keadilan dan kejujuran dalam urusan ekonomi. Konsep ini dapat digunakan dalam konteks segmentasi *Customer Lifetime Value* (CLV) dengan menggunakan algoritma clustering *Expectation Maximization* (EM) dan pendekatan hubungan LRFM dengan memastikan bahwa setiap segmen pelanggan diperlakukan sama dan tidak terjadi eksploitasi. Dunia usaha dapat menggunakan teknik ini untuk menentukan pelanggan bernilai tinggi dan menawarkan layanan yang sesuai, memastikan bahwa semua transaksi dilakukan atas dasar suka sama suka dan didasarkan pada kepercayaan, sesuai dengan prinsip-prinsip yang diuraikan dalam ayat tersebut.

يٰٓاَيُّهَا الَّذِيْنَ اٰمَنُوْا لَا تَأْكُلُوْٓا اَمْوَالَكُمْ بَيْنَكُمْ بِالْبَاطِلِ اِلَّآ اَنْ تَكُوْنَ تِجَارَةً عَنْ تَرَاضٍ مِّنْكُمْۗ وَلَا تَقْتُلُوْٓا اَنْفُسَكُمْۗ اِنَّ اللّٰهَ كَانَ بِكُمْ رَحِيْمًا

Yang artinya: "*Wahai orang-orang yang beriman, janganlah kamu memakan harta sesamamu dengan cara yang batil (tidak benar), kecuali berupa perniagaan atas dasar suka sama suka di antara kamu. Janganlah kamu membunuh dirimu. Sesungguhnya Allah adalah Maha Penyayang kepadamu."* (An-Nisa ayat 29)

Ayat ini melarang pencurian barang milik orang lain dengan cara yang curang, kecuali transaksi atas persetujuan bersama. Para ulama tafsir menjelaskan bahwa larangan ayat tersebut memakan harta orang lain mempunyai makna yang mendalam dan luas yang meliputi:

1. Iman Islam mengakui realitas hak milik pribadi yang tidak dapat disangkal dan berhak dipertahankan.

2. Hak milik pribadi dikenakan pembayaran zakat dan kewajiban lainnya untuk kepentingan negara, agama, dan organisasi lainnya, asalkan memenuhi nisab.
3. Harta seseorang tidak boleh diambil tanpa izin pemiliknya, meskipun harta itu banyak dan dibutuhkan oleh beberapa anggota organisasi yang berhak menerima zakat.

Berdagang atau jual beli untuk mengejar kekayaan boleh saja, asalkan kedua belah pihak bersedia dan tidak ada paksaan. Karena kalaupun ada pembayaran atau penggantian, jual beli yang dipaksakan tidak sah. Tidak boleh ada aspek ketidakadilan terhadap orang lain individu atau masyarakat secara keseluruhan dalam upaya mencapai kesejahteraan. perbuatan yang dilakukan karena keserakahan untuk memperoleh harta benda, seperti suap, riba, perjudian, korupsi, pencurian, dan sebagainya (Arief Mizuary, 2019).

Dalam surat An-Nisa Ayat 29 mengajarkan pentingnya bersikap adil dalam bertransaksi dan menekankan perlunya menjauhi segala bentuk eksploitasi dan penipuan. Perihal menjaga asas keadilan dan integritas dalam hubungan bisnis. Gagasan keadilan diwujudkan dalam penelitian ini dengan menentukan nilai asli setiap pelanggan berdasarkan perilaku transaksional mereka melalui penerapan teknik analisis yang metodis dan obyektif. Hal ini sesuai dengan hikmah surat An-Nisaa 29 yang menekankan pentingnya menyikapi setiap transaksi secara adil dan jujur serta menjunjung tinggi prinsip-prinsip moral yang terkait dengan keadilan dalam segala bidang kehidupan. Termasuk dalam pengelolaan nilai pelanggan dalam perusahaan.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan yang sudah dilakukan sebelumnya, hasil akhir ari klasterisasi pelanggan maskapai berdasarkan model *Length Recency Frequency* dan *Monetary* dengan menggunakan algoritma *expectation maximization* yang berhenti pada maksimal iterasi ke-500. Hasil dari perhitungan algoritma *expectation maximization* mendapatkan jumlah *cluster* yang sama dengan algoritma *K Mean* yang dikelompokkan menjadi 4 *cluster* yaitu *cluster* 0, *cluster* 1, *cluster* 2 dan *cluster* 3. Pada *cluster* 0 dengan nilai *loyal* terdapat 1135 pelanggan, *cluster* 1 dengan nilai *potential* terdapat 2873, *cluster* 2 dengan nilai *general & low value* terdapat 3149 pelanggan serta *cluster* 3 dengan nilai *important* terdapat 731 pelanggan. Keempat *cluster* tersebut memiliki nilai derajat keanggotaan yang terdapat pada algoritma *expectation maximization*. Data hasil *cluster* menggunakan algoritma *expectation maximization* memiliki tingkat kemiripan sebesar 84% dengan perhitungan variasi 15 *principal component analysis* dengan *silhouette scores* sebesar 0,23662.

5.2 Saran

Masih banyak kekurangan dalam pembuatan sistem klaterisasi data pelanggan maskapai dengan algoritma *expectation maximization*. untuk penelitian selanjunya diharapkan:

- Dapat dibangun dengan metode *clustering* lainnya, untuk melihat dan membandingkan keberhasilan klasterisasi data.
- Dapat menambah jumlah data dan atribut serta menggunakan konfigurasi yang berbeda untuk mencapai hasil *cluster* yang lebih maksimal
- Nilai error yang diharapkan bisa dikecilkan lagi, karena semakin kecil nilai error maka akan didapatkan hasil yang lebih baik

DAFTAR PUSTAKA

Agri Ardyan, A., & Budi Darmawan, J. (2016). Sistem pemerolehan informasi karya ilmiah berbasis cluster dengan g-means clustering. In *Seminar Riset Teknologi Informasi (SRITI) tahun*.

Alasadi, S. A., & Bhaya, W. S. (2017). Review of data preprocessing techniques in data mining. *Journal of Engineering and Applied Sciences*, *12*(16), 4102–4107. https://doi.org/10.3923/jeasci.2017.4102.4107

Arief Mizuary. (2019). Fiqih ekonomi Qur'an An-Nisa 29. *Pustaka Pranala*.

Čermák, P. (2015). Customer Profitability Analysis and Customer Life Time Value Models: Portfolio Analysis. *Procedia Economics and Finance*, *25*, 14–25. https://doi.org/10.1016/s2212-5671(15)00708-x

Chakraborty, S., & Das, S. (2018). Simultaneous variable weighting and determining the number of clusters—A weighted Gaussian means algorithm. *Statistics and Probability Letters*, *137*, 148–156. https://doi.org/10.1016/j.spl.2018.01.015

Christy, A. J., Umamakeswari, A., Priyatharsini, L., & Neyaa, A. (2021). RFM ranking – An effective approach to customer segmentation. *Journal of King Saud University - Computer and Information Sciences*, *33*(10), 1251–1257. https://doi.org/10.1016/j.jksuci.2018.09.004

Damuri, A., Riyanto, U., Rusdianto, H., & Aminudin, M. (2021). Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako. *Jurnal Riset Komputer)*, *8*(6), 2407–389. https://doi.org/10.30865/jurikom.v8i6.3655

Ditendra, E., Monalisa, S., Anderjovi, S., & Lesmana, S. (2020a). Klasterisasi clv dengan model lrfm menggunakan algoritma fuzzy c-means (Studi Kasus: Pangeran Gym Pekanbaru) 1. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, *6*(1), 109–116.

Ditendra, E., Monalisa, S., Anderjovi, S., & Lesmana, S. (2020b). Klasterisasi clv dengan model lrfm menggunakan algoritma fuzzy c-means (Studi Kasus: Pangeran Gym Pekanbaru) 1. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, *6*(1), 109–116.

Dyantina, O., Afrina, M., & Ibrahim, A. (2012). Penerapan Customer Relationship Management (CRM) Berbasis Web (Studi Kasus Pada Sistem Informasi Pemasaran di Toko YEN-YEN). *Jurnal Sistem Informasi (JSI)*, *4*(2), 516–529. http://ejournal.unsri.ac.id/index.php/jsi/index

Gesta Nabilla, A., & Tuasela, A. (2021). Strategi pemasaran dalam upaya meningkatkan pendapatan pada diva karaoke rumah bernyanyi di kota timika. *Jurnal kritis*, *5*.

Gopal, S., Patro, K., & Kumar Sahu, K. (n.d.). *Normalization: A Preprocessing Stage*. www.kiplinger.com,

Heri Susanto. (2014). Data mining untuk memprediksi prestasi siswa berdasarkan sosial ekonomi, motivasi, kedisiplinan dan prestasi masa lalu. *Jurnal Vokasi*, *4*(2).

Jolliffe. IT. (2019). *Principal Component Analysis*, *Second Edition*.

Kartika Zahretta Wijaya, Arif Djunaidy, & Faizal Mahananto. (2021). Segmentasi Pelanggan Menggunakan Algoritma K-Means dan Analisis RFM di Ova Gaming E-Sports Arena Kediri. *Jurnal Teknik Its*, *10*.

Marisa Efendi, D., Sartika, D., Isnayah Waspah, A., Afandi, A., Informasi, S., & Dian Cipta Cendikia Kotabumi, S. (2022). Expectation maximization algorithm memprediksi penjualan susu murni pada pt. Sewu primatama indonesia lampung tengah. In *Jurnal Teknik Informatika Musirawas) Aik Isnayah Waspah, Asep Afandi* (Vol. 7, Issue 1).

Moh rusdi. (2019). Strategi Pemasaran Untuk Meningkatkan Volume Penjualan Pada Perusahaan Genting Ud. Berkah Jaya. *Jurnal Studi Manajemen Dan Bisnis*, *6*(2), 49–54. http://journal.trunojoyo.ac.id/jsmb

Monalisa, S. (2018a). Klusterisasi Customer Lifetime Value dengan Model LRFM menggunakan Algoritma K-Means. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *5*(2), 247–252. https://doi.org/10.25126/jtiik.201852690

Monalisa, S. (2018b). Klusterisasi Customer Lifetime Value dengan Model LRFM menggunakan Algoritma K-Means. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *5*(2), 247. https://doi.org/10.25126/jtiik.201852690

Nur Atikah, Swasono Rahardjo, Trianingsih Eni L, & Lucky Tri Oktoviana. (2021). Penerapan algoritma expectation-maximization (em) dalam mengelompokkan popularitas objek wisata di malang raya berdasarkan indikator banyak pengunjung. *Jurnal Kajian Matematika Dan Aplikasinya*, *2*.

Parvaneh, A., Abbasimehr, H., & Tarokh, M. J. (2012). Integrating AHP and Data Mining for Effective Retailer Segmentation Based on Retailer Lifetime Value. *Journal of Optimization in Industrial Engineering*.

Pratomo, edwin agung, Najib, M., & Mulyati, H. (2019). Customer segmentation analysis based on the customer lifetime value method. *Jurnal Aplikasi Manajemen*, *17*(3), 408–415. https://doi.org/10.21776/ub.jam.2019.017.03.04

Rosmayani. (2016). Customer relationship management. *Jurnal Valuta*, *2*(1), 83–98.

Sirait, R. E., Darwianto, E., Dwi, D., & Suwawi, J. (n.d.). *Implementasi dan AnalisisAlgoritma Clustering Expectation-maximization (EM) Pada Data Tugas Akhir Universitas Telkom*.

Smith, G. (2020). Data mining fool's gold. *Journal of Information Technology*, *35*(3), 182–194. https://doi.org/10.1177/0268396220915600