

Penerapan Algoritma K-Means Clustering untuk Segmentasi Pelanggan Berdasarkan Model RFM pada Dataset Online Retail II

Rendi Suoth¹, Veto Almalik², Rendi Ramadhani³

¹²³ Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pamulang, Tangerang Selatan, Indonesia

¹ rendisuoth23@gmail.com, ² vetoalmalik7@gmail.com, ³ ren.rahmadani@gmail.com

Abstract

The advancement of information technology has encouraged companies to utilize customer transaction data as a valuable source of insight to support business decision-making. One approach that can be applied is customer segmentation, which enables companies to understand customer characteristics and purchasing behavior based on their transaction activities. This study aims to implement the K-Means Clustering algorithm based on the RFM (Recency, Frequency, Monetary) model to segment customers using the Online Retail II dataset. The research adopts a quantitative approach through data mining stages, including business understanding, data understanding, data preparation, modeling, evaluation, and deployment. During the data preparation stage, data cleaning, missing value removal, invalid transaction filtering, and RFM feature construction were performed. Subsequently, the K-Means algorithm was applied to group customers according to their transaction characteristics. The results indicate that the optimal number of clusters determined using the Elbow Method is three, consisting of loyal customers, regular customers, and inactive customers. The loyal customer cluster is characterized by low Recency values and very high Frequency and Monetary values, whereas the inactive customer cluster exhibits high Recency values with low Frequency and Monetary values. Furthermore, this study developed a customer intelligence dashboard that presents customer segmentation, churn-risk customers, product bundling recommendations, and transaction value predictions. The findings demonstrate that the integration of the RFM model and K-Means Clustering algorithm can produce meaningful customer segments and serve as a foundation for developing more effective and targeted marketing strategies.

Keywords: : Data Mining, Customer Segmentation, RFM, K-Means Clustering, Online Retail II, Customer Intelligence Dashboard.

Abstrak

Kemajuan teknologi informasi telah mendorong perusahaan untuk memakai data transaksi konsumen sebagai sumber wawasan penting untuk mendukung keputusan bisnis. Salah satu cara yang bisa dilakukan adalah dengan membagi pelanggan ke dalam kelompok-kelompok agar bisa memahami ciri-ciri dan kebiasaan konsumen berdasarkan kegiatan transaksi mereka. Tujuan dari kajian ini adalah mengaplikasikan metode K-Means Clustering berdasarkan model RFM (Recency, Frequency, Monetary) untuk mengelompokkan konsumen pada kumpulan data Online Retail II. Pendekatan yang dipakai dalam kajian ini adalah kuantitatif melalui langkah-langkah data mining yang meliputi pemahaman bisnis, pemahaman data, penyiapan data, pemodelan, evaluasi, dan implementasi. Dalam tahap penyiapan data, dilakukan pembersihan data, penyingkiran nilai yang hilang, penghapusan transaksi yang tidak sah, serta pembentukan fitur RFM. Selanjutnya, metode K-Means diterapkan untuk menggolongkan konsumen berdasarkan ciri transaksi yang mereka miliki. Hasil dari kajian ini memperlihatkan bahwa jumlah kelompok terbaik yang ditemukan melalui Elbow Method adalah tiga kelompok, yaitu konsumen setia, konsumen umum, dan konsumen tidak aktif. Kelompok konsumen setia memiliki nilai Recency yang rendah, serta nilai Frequency dan Monetary yang sangat tinggi, sementara kelompok konsumen tidak aktif ditandai dengan nilai Recency yang tinggi dan nilai Frequency serta Monetary yang rendah. Lebih jauh, kajian ini juga menciptakan sebuah dashboard intelijen konsumen yang menyajikan pengelompokan konsumen, konsumen yang berpotensi berhenti, saran paket produk, dan perkiraan nilai transaksi. Temuan penelitian menunjukkan bahwa gabungan model RFM dan metode K-Means dapat menghasilkan pengelompokan konsumen yang jelas dan bisa menjadi dasar dalam menyusun strategi pemasaran yang lebih baik dan tertarget.

Kata Kunci: : Data Mining, Segmentasi Pelanggan, RFM, K-Means Clustering, Online Retail II, Customer Intelligence Dashboard

A. PENDAHULUAN

Perkembangan teknologi informasi dan transformasi

digital telah membawa perubahan yang signifikan pada berbagai sektor industri, khususnya dalam pengelolaan data pelanggan dan transaksi bisnis. Perusahaan saat ini menghasilkan data transaksi dalam jumlah yang sangat besar setiap harinya, terutama pada sektor retail online yang terus mengalami pertumbuhan. Data transaksi tersebut tidak hanya berfungsi sebagai arsip operasional, tetapi juga dapat dimanfaatkan sebagai sumber informasi yang bernilai dalam mendukung pengambilan keputusan bisnis. Menurut Dharma dan Fenriana (2024), pemanfaatan data pelanggan secara optimal dapat membantu perusahaan memahami karakteristik pelanggan sehingga strategi pemasaran dapat dilakukan secara lebih efektif dan terarah. Selain itu, meningkatnya persaingan bisnis digital menuntut perusahaan untuk mampu mengelola dan menganalisis data pelanggan secara lebih sistematis guna mempertahankan loyalitas pelanggan serta meningkatkan keuntungan perusahaan. Oleh karena itu, pengelolaan data pelanggan menjadi salah satu aspek penting dalam pengembangan sistem informasi bisnis modern.

Dalam pengelolaan bisnis berbasis digital, data mining memiliki peran penting dalam mengekstraksi pola dan informasi tersembunyi dari kumpulan data yang berukuran besar. Melalui teknik data mining, perusahaan dapat menemukan hubungan antar data, pola perilaku pelanggan, serta karakteristik transaksi yang sebelumnya sulit diidentifikasi secara manual. Menurut Fadhilah, Suyoso, dan Puspitasari (2024), penerapan data mining mampu membantu perusahaan dalam melakukan analisis pelanggan secara lebih cepat, akurat, dan berbasis data. Salah satu implementasi data mining yang banyak digunakan dalam bidang pemasaran adalah customer segmentation atau segmentasi pelanggan. Teknik ini digunakan untuk mengelompokkan pelanggan berdasarkan karakteristik tertentu sehingga perusahaan dapat menentukan strategi pemasaran yang lebih sesuai dengan kebutuhan setiap kelompok pelanggan. Dengan adanya segmentasi pelanggan, perusahaan dapat meningkatkan efektivitas promosi, retensi pelanggan, dan kualitas pelayanan kepada pelanggan.

Customer segmentation merupakan strategi bisnis yang bertujuan mengelompokkan pelanggan berdasarkan kesamaan perilaku maupun karakteristik transaksi yang dimiliki. Setiap pelanggan memiliki pola pembelian, tingkat loyalitas, dan kontribusi nilai transaksi yang berbeda sehingga memerlukan pendekatan pemasaran yang berbeda pula. Menurut Tivano, Ruslianto, dan Prawira (2023), segmentasi pelanggan membantu perusahaan mengidentifikasi kelompok pelanggan potensial serta menentukan strategi bisnis yang lebih tepat sasaran. Melalui segmentasi pelanggan, perusahaan dapat mengetahui pelanggan yang aktif, pelanggan loyal, maupun pelanggan yang berpotensi berhenti melakukan transaksi. Informasi tersebut sangat penting dalam mendukung pengambilan keputusan terkait promosi, layanan pelanggan, dan pengelolaan hubungan pelanggan. Dengan demikian, segmentasi pelanggan dapat membantu perusahaan memaksimalkan keuntungan sekaligus

menjaga hubungan jangka panjang dengan pelanggan.

Penelitian ini menggunakan dataset Online Retail II sebagai objek penelitian. Dataset tersebut berisi data transaksi pelanggan retail online yang mencakup informasi nomor transaksi, produk yang dibeli, jumlah pembelian, harga produk, tanggal transaksi, serta identitas pelanggan. Dataset Online Retail II banyak digunakan dalam penelitian segmentasi pelanggan karena memiliki jumlah data yang besar dan mampu merepresentasikan aktivitas transaksi pelanggan secara nyata. Menurut Nurhidayat, Kusuma, dan Septiawan (2025), dataset Online Retail II sangat sesuai digunakan dalam penelitian customer segmentation karena memiliki atribut transaksi yang lengkap untuk membangun model RFM. Data transaksi pada dataset tersebut memungkinkan proses analisis perilaku pelanggan dilakukan secara lebih komprehensif berdasarkan aktivitas pembelian yang dilakukan pelanggan. Oleh karena itu, dataset ini dinilai relevan untuk digunakan dalam penelitian segmentasi pelanggan berbasis data mining.

Dalam pengelolaan data pelanggan, perusahaan sering mengalami kesulitan dalam mengidentifikasi pelanggan loyal, pelanggan pasif, maupun pelanggan yang memiliki potensi tinggi terhadap peningkatan penjualan. Banyak perusahaan masih melakukan analisis pelanggan secara manual sehingga proses identifikasi pelanggan menjadi kurang efektif dan membutuhkan waktu yang relatif lama. Selain itu, perusahaan juga sering mengalami kesulitan dalam menentukan strategi pemasaran yang tepat karena karakteristik pelanggan yang beragam. Ketidakmampuan dalam memahami perilaku pelanggan dapat menyebabkan promosi tidak tepat sasaran sehingga biaya pemasaran menjadi kurang efisien. Menurut Passolving, Ali, Kartawidjaja, dan Sukwadi (2024), kurangnya pemahaman terhadap karakteristik pelanggan dapat mengurangi efektivitas strategi pemasaran dan menghambat upaya peningkatan loyalitas pelanggan. Kondisi tersebut menunjukkan pentingnya penerapan metode analisis pelanggan yang lebih sistematis dan berbasis data.

Permasalahan dalam pengelolaan pelanggan dapat memberikan dampak signifikan terhadap efektivitas strategi bisnis perusahaan. Strategi promosi yang tidak sesuai dengan karakteristik pelanggan dapat menyebabkan rendahnya tingkat keberhasilan pemasaran serta pemborosan biaya operasional. Selain itu, perusahaan berisiko kehilangan pelanggan bernilai tinggi apabila tidak memiliki sistem yang mampu mengidentifikasi pelanggan potensial secara tepat. Dalam jangka panjang, kondisi tersebut dapat menurunkan daya saing perusahaan di tengah meningkatnya persaingan bisnis digital. Menurut Tivano et al. (2023), segmentasi pelanggan yang dilakukan secara tepat dapat membantu perusahaan menentukan prioritas pelanggan dan meningkatkan efektivitas strategi pemasaran. Oleh sebab itu, diperlukan suatu metode yang mampu membantu perusahaan melakukan segmentasi pelanggan secara otomatis, cepat, dan akurat. Untuk mengatasi permasalahan tersebut, penelitian ini mengaplikasikan model RFM (Recency, Frequency,

Monetary) serta algoritma K-Means Clustering dalam membagi pelanggan menjadi beberapa segmen. Model RFM bertugas menganalisis kebiasaan belanja pelanggan berdasarkan kapan terakhir mereka bertransaksi, seberapa sering mereka bertransaksi, dan jumlah total uang yang mereka belanjakan. Fadhillah dkk. (2024) menjelaskan bahwa model RFM merupakan cara efektif untuk mengukur seberapa terlibat dan setia pelanggan didasarkan pada catatan transaksi mereka. Selanjutnya, algoritma K-Means Clustering dipakai untuk mengelompokkan pelanggan berdasarkan kesamaan atribut RFM yang mereka miliki. Lukas, Okmayura, Irani, dan kolega (2025) meyakini bahwa perpaduan model RFM dan algoritma K-Means dapat menghasilkan segmentasi pelanggan yang lebih tertata dan gampang dicerna oleh bisnis. Hasil pengelompokan ini lantas dipakai untuk mengidentifikasi pelanggan yang loyal, pelanggan yang cukup sering berbelanja, dan pelanggan yang jarang berbelanja, agar perusahaan bisa merancang taktik bisnis yang sesuai untuk tiap golongan pelanggan.

Selain melakukan segmentasi pelanggan, penelitian ini juga mengembangkan visualisasi dashboard customer intelligence untuk membantu perusahaan memahami hasil analisis secara lebih interaktif. Dashboard yang dibuat menampilkan informasi mengenai segmentasi pelanggan, pelanggan berisiko churn, rekomendasi bundling produk, serta prediksi nilai transaksi pelanggan. Menurut Dharma dan Fenriana (2024), visualisasi hasil analisis dalam bentuk dashboard interaktif dapat membantu pengguna memahami informasi secara lebih cepat dan mendukung pengambilan keputusan yang lebih efektif. Dashboard juga memungkinkan perusahaan memantau perilaku pelanggan secara lebih mudah melalui penyajian informasi yang ringkas dan informatif. Integrasi antara data mining dan dashboard customer intelligence diharapkan mampu meningkatkan efektivitas sistem informasi perusahaan dalam mendukung pengelolaan hubungan pelanggan.

Berdasarkan uraian tersebut, tujuan penelitian ini adalah menerapkan algoritma K-Means Clustering berdasarkan model RFM untuk melakukan segmentasi pelanggan pada dataset Online Retail II. Penelitian ini juga bertujuan membantu perusahaan dalam mengidentifikasi pelanggan loyal, pelanggan pasif, serta pelanggan yang berpotensi memberikan keuntungan tinggi bagi perusahaan. Selain itu, penelitian ini bertujuan mengembangkan dashboard customer intelligence yang mampu menampilkan hasil segmentasi pelanggan secara interaktif sehingga mendukung pengambilan keputusan bisnis. Dengan adanya penelitian ini, diharapkan perusahaan dapat meningkatkan efektivitas strategi pemasaran, mempertahankan loyalitas pelanggan, serta meningkatkan penjualan melalui pemanfaatan teknologi data mining dan sistem informasi secara optimal.

B. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan teknik penambangan data guna menilai tingkah laku

pembeli dari catatan transaksi ritel online. Pendekatan kuantitatif dipilih karena penelitian ini berfokus pada analisis data numerik serta pola transaksi pembeli dengan memakai algoritma pengelompokan. Teknik penambangan data dimanfaatkan untuk menemukan pola tersembunyi dalam data transaksi pembeli, yang dapat menyederhanakan pembagian pelanggan secara otomatis. Menurut Han, Kamber, dan Pei (2022), penambangan data merupakan sebuah proses penggalian informasi guna menemukan pola, kaitan, dan pengetahuan bermanfaat dari kumpulan data yang luas. Penelitian ini bertujuan untuk mengelompokkan pembeli berdasarkan ciri tingkah laku pembelian dengan memakai model RFM dan algoritma K-Means Clustering. Diharapkan hasil pembagian pelanggan ini dapat memberikan pemahaman bagi perusahaan untuk mengenal ciri pembeli dan mendukung penentuan bisnis yang lebih baik.

Objek penelitian yang digunakan adalah dataset Online Retail II yang diperoleh dari Kaggle sebagai sumber dataset publik. Dataset tersebut berisi data transaksi pelanggan retail online yang mencakup informasi nomor transaksi, deskripsi produk, jumlah pembelian, harga produk, tanggal transaksi, dan identitas pelanggan. Dataset Online Retail II dipilih karena memiliki jumlah data transaksi yang besar dan sering digunakan dalam penelitian customer segmentation berbasis data mining. Teknik pengumpulan data dalam penelitian ini dilakukan melalui studi pustaka dan pengambilan dataset publik. Studi pustaka dilakukan dengan mempelajari jurnal ilmiah, artikel penelitian, serta referensi terkait metode RFM, K-Means Clustering, customer segmentation, dan dashboard customer intelligence. Menurut Fadhillah, Suyoso, dan Puspitasari (2024), penggunaan dataset transaksi retail memberikan gambaran yang representatif dalam menganalisis perilaku pelanggan melalui pendekatan RFM dan clustering.

Tahapan penelitian dilakukan berdasarkan alur proses data mining yang terdiri dari business understanding, data understanding, data preparation, modeling, evaluation, dan deployment. Model tersebut mengacu pada metodologi CRISPDM yang banyak digunakan dalam proyek data mining. Menurut Wirth dan Hipp (2000), CRISP-DM menyediakan kerangka kerja sistematis yang membantu peneliti dalam mengelola proses analisis data secara terstruktur. Pada tahap business understanding dilakukan identifikasi permasalahan bisnis terkait kesulitan perusahaan dalam memahami perilaku pelanggan dan menentukan strategi pemasaran yang tepat. Selanjutnya dilakukan data understanding dan data preparation melalui proses preprocessing data yang meliputi pembersihan data, penghapusan missing value, filtering transaksi tidak valid, serta transformasi data menjadi atribut RFM. Tahapan ini bertujuan menghasilkan data yang lebih bersih dan siap digunakan pada proses pemodelan.

Model RFM (Recency, Frequency, Monetary) digunakan dalam penelitian ini untuk menganalisis perilaku pelanggan berdasarkan aktivitas transaksi. Recency digunakan untuk mengukur jarak waktu sejak pelanggan terakhir melakukan

transaksi, Frequency digunakan untuk menghitung jumlah transaksi pelanggan, sedangkan Monetary digunakan untuk menghitung total nilai pembelian pelanggan. Menurut Tivano, Ruslianto, dan Prawira (2023), model RFM merupakan salah satu metode yang efektif dalam mengidentifikasi tingkat loyalitas pelanggan berdasarkan riwayat transaksi yang dimiliki. Setelah nilai RFM diperoleh, data dinormalisasi agar setiap atribut memiliki skala yang seimbang sebelum dilakukan proses clustering. Menurut Lukas et al. (2025), normalisasi data penting dilakukan untuk mengurangi pengaruh perbedaan skala antar atribut sehingga hasil clustering menjadi lebih optimal. Dengan menggunakan model RFM, perusahaan dapat memperoleh gambaran perilaku pelanggan secara lebih akurat berdasarkan aktivitas transaksi yang dilakukan.

Proses mengelompokkan pelanggan memakai algoritma K-Means Clustering yang bertujuan membuat kelompok berdasarkan kesamaan pada karakteristik RFM. Algoritma K-Means dipilih karena kemampuannya yang hebat dalam mengelompokkan data jumlah besar secara cepat serta mudah. Seperti dijelaskan oleh Tan, Steinbach, Karpatne, dan Kumar (2019), K-Means adalah salah satu algoritma pengelompokan yang paling sering dipakai sebab efisiensinya dalam mengelompokkan objek berdasarkan jarak ke centroid cluster. Jumlah cluster ditentukan memakai Metode Elbow dengan menganalisis nilai inerti pada beberapa jumlah cluster yang berbeda. Setelah tahap pengelompokan selesai dilakukan, tiap cluster dianalisis berdasarkan nilai rata-rata RFM guna mengenali karakteristik dari masing-masing kelompok pelanggan. Hasil pengelompokan tersebut kemudian diartikan menjadi beberapa kategori pelanggan seperti pelanggan setia, pelanggan biasa, dan pelanggan pasif.

Visualisasi hasil clustering dilakukan menggunakan grafik dan dashboard customer intelligence berbasis Streamlit. Dashboard dikembangkan untuk menampilkan hasil segmentasi pelanggan secara interaktif sehingga memudahkan pengguna dalam memahami hasil analisis data mining. Dashboard yang dibuat menampilkan distribusi segmentasi pelanggan, pelanggan berisiko churn, rekomendasi bundling produk, serta prediksi nilai transaksi pelanggan. Rekomendasi bundling produk dihasilkan menggunakan metode Apriori dan Association Rules untuk menemukan pola pembelian produk yang sering dibeli secara bersamaan. Menurut Hahsler (2023), Association Rules Mining dapat digunakan untuk menemukan hubungan antarproduk yang sering muncul dalam transaksi pelanggan. Selain itu, prediksi nilai transaksi dilakukan menggunakan metode Linear Regression berdasarkan nilai frequency pelanggan.

Tahap evaluasi dilakukan dengan menganalisis hasil clustering berdasarkan karakteristik masing-masing cluster serta interpretasi perilaku pelanggan. Evaluasi hasil clustering dilakukan melalui analisis rata-rata nilai RFM pada setiap cluster untuk mengetahui tingkat loyalitas dan aktivitas pelanggan. Menurut Passolving et al. (2024), keberhasilan segmentasi pelanggan dapat diukur dari

kemampuan cluster dalam menunjukkan perbedaan karakteristik yang jelas antar kelompok pelanggan. Teknologi dan tools yang digunakan dalam penelitian ini meliputi Python sebagai bahasa pemrograman utama, Pandas untuk pengolahan data, Scikit-Learn untuk implementasi algoritma machine learning, Matplotlib dan Seaborn untuk visualisasi data, serta Streamlit untuk pengembangan dashboard interaktif. Dengan tahapan penelitian tersebut, diharapkan sistem segmentasi pelanggan yang dihasilkan mampu membantu perusahaan dalam memahami perilaku pelanggan dan meningkatkan efektivitas strategi pemasaran.

C. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk mengidentifikasi karakteristik pelanggan berdasarkan perilaku transaksi menggunakan model RFM (Recency, Frequency, Monetary) dan algoritma K-Means Clustering pada dataset Online Retail II. Dataset yang digunakan berisi data transaksi pelanggan retail online yang mencakup informasi waktu transaksi, jumlah pembelian, harga produk, dan identitas pelanggan. Sebelum proses clustering dilakukan, data terlebih dahulu melalui tahapan preprocessing untuk meningkatkan kualitas data yang akan dianalisis. Tahapan tersebut meliputi pembersihan data, penghapusan nilai yang hilang (missing value), penghapusan transaksi yang tidak valid, serta transformasi data menjadi atribut RFM. Hasil preprocessing menunjukkan bahwa data yang digunakan telah memenuhi kebutuhan analisis dan mampu merepresentasikan aktivitas transaksi pelanggan secara lebih akurat. Menurut Fadhillah, Suyoso, dan Puspitasari (2024), kualitas data yang baik merupakan faktor penting dalam menghasilkan segmentasi pelanggan yang akurat karena proses clustering sangat bergantung pada kualitas data yang digunakan sebagai input.

Model RFM digunakan dalam penelitian ini karena efektif dalam menggambarkan kebiasaan pembeli berdasarkan tiga bagian pokok, yaitu Keterkinian, Frekuensi, dan Nilai. Keterkinian berfungsi untuk mengukur sudah berapa lama pelanggan tidak bertransaksi, Frekuensi menghitung seberapa sering pelanggan melakukan transaksi, sementara Nilai mengukur total uang yang dibelanjakan oleh pelanggan. Ketiga bagian ini memberikan gambaran lengkap mengenai seberapa terhubungnya pelanggan dengan bisnis. Pelanggan dengan nilai Keterkinian rendah menunjukkan transaksi yang lebih baru dibandingkan dengan mereka yang memiliki nilai Keterkinian tinggi. Di sisi lain, nilai Frekuensi dan Nilai yang tinggi mencerminkan kesetiaan dan kontribusi pendapatan yang lebih besar. Menurut Tivano, Ruslianto, dan Prawira (2023), gabungan dari karakteristik RFM dapat memberikan data yang lebih rinci dalam pengelompokan pelanggan dibandingkan dengan hanya memakai satu penanda transaksi..

Setelah menghitung nilai RFM untuk setiap pelanggan, pengelompokan dilakukan menggunakan algoritma KMeans. Algoritma ini dipilih karena kemampuannya dalam mengelompokkan data berdasarkan kesamaan

karakteristik pelanggan. Sebelum dikelompokkan, data RFM perlu dinormalisasi agar setiap atribut memiliki skala yang sama dan mencegah dominasi atribut tertentu dalam pembentukan kluster. Hasil normalisasi menunjukkan bahwa semua atribut dapat dimanfaatkan sepenuhnya dalam proses pengelompokan. Menurut Lukas et al. (2025), normalisasi data merupakan langkah penting dalam menerapkan algoritma K-Means, karena perbedaan rentang nilai antar atribut dapat memengaruhi posisi centroid dan kualitas kluster. Oleh karena itu, normalisasi dilakukan untuk meningkatkan kualitas segmentasi pelanggan.

Berdasarkan hasil analisis menggunakan Metode Elbow, jumlah kluster paling ideal ditemukan adalah tiga kluster. Penentuan jumlah kluster dilakukan dengan memeriksa penurunan nilai inertia pada setiap nilai k yang diuji. Visualisasi hasil menunjukkan bahwa penurunan inertia mulai stabil saat $k = 3$, sehingga angka tersebut dipilih sebagai jumlah kluster paling akurat untuk mencerminkan karakteristik pelanggan. Melalui penggunaan tiga kluster, pelanggan dapat dikelompokkan ke dalam kategori yang menunjukkan pola transaksi yang cukup berbeda. Menurut penelitian Passolving et al. (2024), jumlah kluster yang optimal akan menciptakan kelompok pelanggan yang seragam dalam satu kluster dan berbeda di antara kluster, sehingga mempermudah pemahaman hasil segmentasi. Temuan yang didapat menunjukkan bahwa algoritma K-Means mampu menghasilkan segmentasi pelanggan yang sesuai dengan karakteristik dari data yang sedang dianalisis.

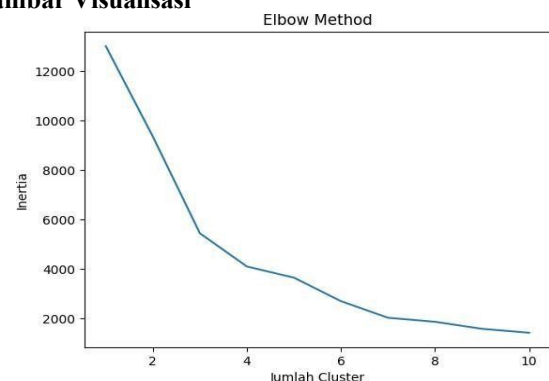
Analisis mengenai karakteristik kelompok menunjukkan adanya perbedaan yang cukup mencolok antara tiga segmen pelanggan yang terbentuk. Segmen pertama memiliki nilai Recency yang relatif rendah, sementara nilai Frequency dan Monetarynya cukup baik, menunjukkan bahwa pelanggan ini masih aktif melakukan pembelian. Segmen kedua memiliki nilai Recency yang tinggi, bersama dengan nilai Frequency dan Monetary yang rendah, yang menandakan bahwa pelanggan ini telah lama tidak melakukan pembelian dan berisiko mengalami churn. Di sisi lain, segmen ketiga menunjukkan nilai Recency yang sangat rendah serta nilai Frequency dan Monetary yang sangat tinggi, sehingga bisa dianggap sebagai kelompok pelanggan yang paling setia dan memberikan kontribusi pendapatan terbesar bagi perusahaan. Perbedaan karakteristik ini menandakan bahwa setiap segmen pelanggan menunjukkan pola transaksi yang unik. Penelitian oleh Nurhidayat, Kusuma, dan Septiawan (2025) menyatakan bahwa hasil segmentasi pelanggan dapat digunakan sebagai fondasi dalam merancang strategi pemasaran yang lebih terarah sesuai dengan karakteristik masing-masing segmen pelanggan.

Hasil segmentasi yang diperoleh memberikan informasi yang penting bagi perusahaan dalam memahami perilaku pelanggan secara lebih mendalam. Melalui hasil clustering, perusahaan dapat mengidentifikasi pelanggan loyal yang perlu dipertahankan, pelanggan biasa yang masih berpotensi untuk ditingkatkan nilai transaksinya, serta

pelanggan pasif yang memerlukan strategi reaktivasi. Informasi tersebut dapat digunakan sebagai dasar dalam menyusun program promosi yang lebih tepat sasaran dan efisien. Selain itu, hasil segmentasi juga membantu perusahaan dalam mengalokasikan sumber daya pemasaran secara lebih efektif kepada kelompok pelanggan yang memiliki nilai bisnis terbesar. Menurut Dharma dan Fenriana (2024), penerapan segmentasi pelanggan berbasis data mining mampu meningkatkan efektivitas strategi pemasaran karena keputusan yang diambil didasarkan pada pola transaksi pelanggan yang nyata. Dengan demikian, kombinasi model RFM dan algoritma K-Means terbukti mampu memberikan gambaran yang jelas mengenai karakteristik pelanggan pada dataset Online Retail II.

Secara keseluruhan, hasil penelitian menunjukkan bahwa penerapan model RFM dan algoritma K-Means Clustering berhasil mengelompokkan pelanggan ke dalam beberapa segmen yang memiliki karakteristik transaksi yang berbeda. Segmentasi yang dihasilkan tidak hanya memberikan informasi mengenai tingkat loyalitas pelanggan, tetapi juga membantu perusahaan dalam mengidentifikasi peluang peningkatan pendapatan dan risiko kehilangan pelanggan. Perbedaan nilai Recency, Frequency, dan Monetary yang terlihat pada setiap cluster menunjukkan bahwa pelanggan memiliki pola perilaku transaksi yang beragam. Hasil tersebut membuktikan bahwa pendekatan data mining dapat dimanfaatkan sebagai alat pendukung pengambilan keputusan dalam pengelolaan hubungan pelanggan. Selain itu, hasil segmentasi yang diperoleh menjadi dasar dalam pengembangan dashboard customer intelligence yang mampu menyajikan informasi pelanggan secara lebih informatif dan mudah dipahami. Dengan demikian, penelitian ini memberikan kontribusi dalam pemanfaatan teknik data mining untuk mendukung strategi bisnis berbasis data pada sektor retail online.

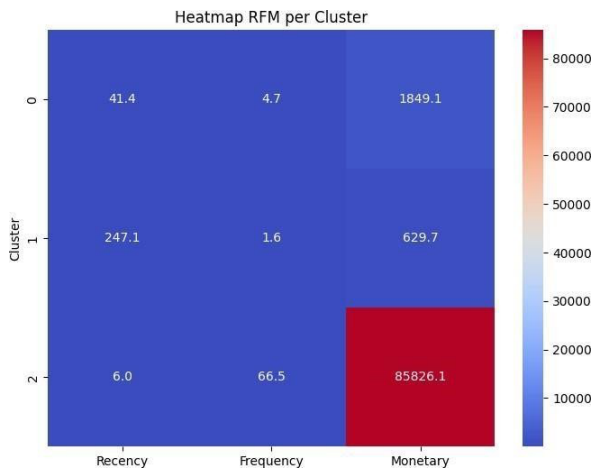
Gambar Visualisasi



Gambar 1 Elbow Method

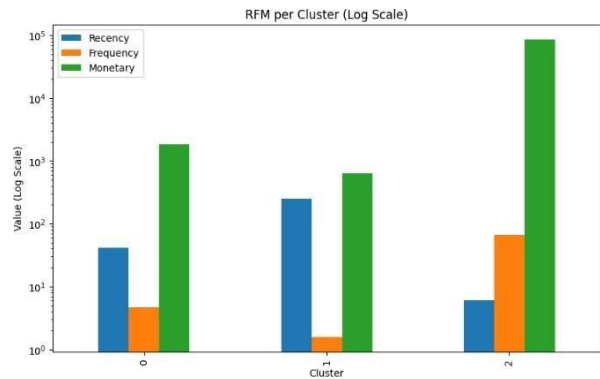
Gambar Elbow Method digunakan untuk menentukan jumlah cluster optimal pada algoritma K-Means Clustering. Grafik menunjukkan hubungan antara jumlah cluster (K) dengan nilai inertia yang dihasilkan. Nilai inertia menggambarkan tingkat kedekatan data terhadap centroid dalam setiap cluster. Semakin kecil nilai inertia menunjukkan bahwa anggota dalam cluster semakin

homogen. Berdasarkan grafik, terjadi penurunan inertia yang sangat signifikan dari K=1 hingga K=3. Setelah K=3, penurunan inertia mulai melandai dan tidak lagi menunjukkan perubahan yang besar. Kondisi ini membentuk pola siku (elbow) pada titik K=3 sehingga dapat disimpulkan bahwa jumlah cluster yang paling optimal adalah tiga cluster. Titik elbow menunjukkan jumlah cluster yang mampu memberikan keseimbangan antara kualitas segmentasi dan kompleksitas model. Oleh karena itu, penelitian ini menggunakan tiga cluster untuk proses segmentasi pelanggan.



Gambar 2 Heatmap RFM per Cluster

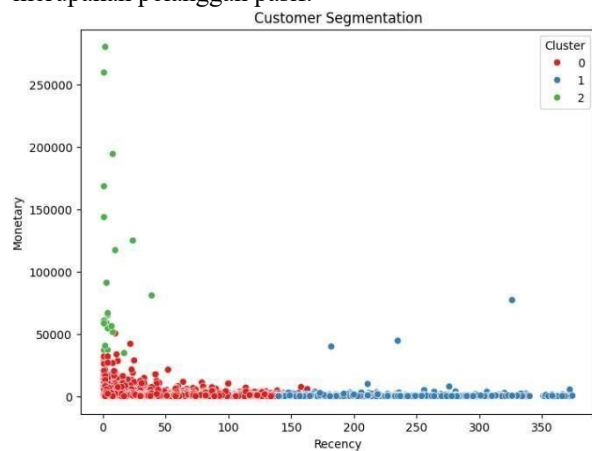
Berdasarkan penguraian nilai RFM rata-rata di tiap grup, kami menemukan perbedaan karakteristik yang jelas. Di Kelompok 0, rata-rata nilai Recency adalah 41,4 hari, Frequency berjumlah 4,7 transaksi, dan Monetary mencapai 1.849,1. Informasi ini mengindikasikan bahwa pelanggan di grup ini masih lumayan sering bertransaksi dengan frekuensi pembelian yang masuk akal. Kelompok 1 mencatat rata-rata nilai Recency sebanyak 247,1 hari, Frequency sebanyak 1,6 transaksi, dan Monetary sebanyak 629,7. Angka Recency yang panjang menunjukkan bahwa para pembeli di grup ini sudah lama tidak melakukan pembelian. Ditambah lagi, nilai Frequency dan Monetary yang rendah menunjukkan rendahnya jumlah aktivitas belanja dari para pembeli di kelompok ini. Kelompok 2 memperlihatkan ciri yang sangat berlainan dibanding kelompok lain. Rata-rata nilai Recency hanya 6 hari, Frequency mencapai 66,5 transaksi, dan Monetary menyentuh 85.826,1. Angka-angka ini mengindikasikan bahwa para pembeli di grup ini sangat giat dalam bertransaksi serta menyumbangkan pemasukan terbesar bagi perusahaan. Oleh sebab itu, Kelompok 2 dapat dianggap sebagai grup pembeli yang loyal.



Gambar 3 RFM per Cluster (Log Scale)

Grafik RFM per Cluster menggunakan skala logaritmik digunakan untuk memperjelas perbedaan nilai antar cluster yang memiliki rentang data sangat jauh. Pada grafik terlihat bahwa nilai Monetary Cluster 2 jauh lebih tinggi dibandingkan Cluster 0 dan Cluster 1. Perbedaan ini menunjukkan bahwa sebagian kecil pelanggan memberikan kontribusi transaksi yang sangat besar terhadap perusahaan. Selain itu, Cluster 2 juga memiliki nilai Frequency tertinggi yang menunjukkan intensitas transaksi yang sangat tinggi.

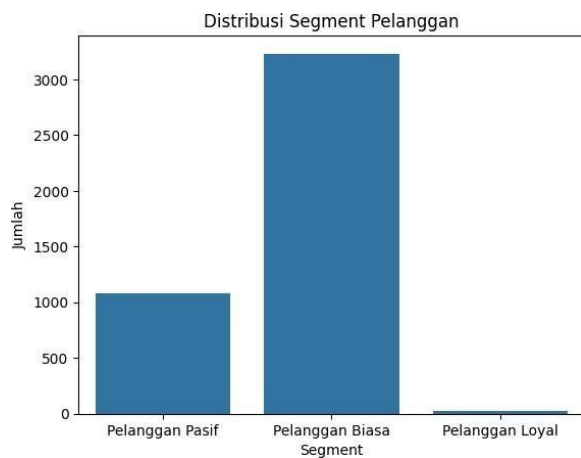
Sebaliknya, Cluster 1 memiliki nilai Recency tertinggi sehingga menunjukkan pelanggan yang sudah lama tidak melakukan pembelian. Hasil visualisasi ini memperkuat interpretasi bahwa Cluster 2 merupakan pelanggan loyal, Cluster 0 merupakan pelanggan biasa, dan Cluster 1 merupakan pelanggan pasif.



Gambar 4 Customer Segmentation

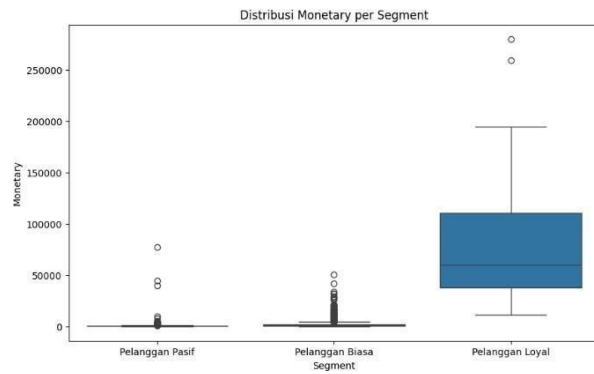
(Sumbu X = Keterkinian, Sumbu Y = Keuangan)
Visualisasi Segmentasi Pelanggan memperlihatkan hasil pengelompokan klien berdasarkan nilai Keterkinian dan Keuangan sebelum penetapan label segmen dilakukan. Pada diagram, Cluster 2 terlihat berfokus pada area dengan nilai Keterkinian rendah dan Keuangan tinggi. Karakteristik ini menunjukkan klien yang baru saja melakukan transaksi dengan nilai pembelian yang signifikan. Di sisi lain, Cluster

1 banyak terletak di area dengan nilai Keterkinian tinggi dan Keuangan rendah. Situasi ini menggambarkan klien yang sudah lama tidak bertransaksi dan memiliki kontribusi pendapatan yang minim. Di samping itu, Cluster 0 berada di antara kedua cluster tersebut dengan karakteristik transaksi yang lebih seimbang. Pemisahan antara titik data yang ada dalam grafik mengindikasikan bahwa algoritma KMeans mampu mengelompokkan klien berdasarkan pola transaksi yang bervariasi. Temuan ini menunjukkan bahwa model yang diterapkan dapat mengenali kelompok klien dengan karakteristik yang jelas.



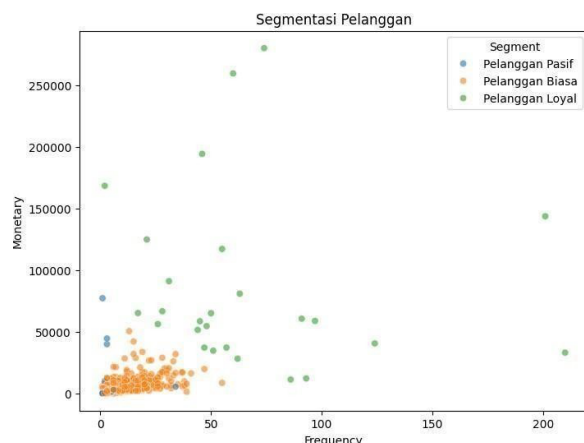
Gambar 5 Distribusi Segment Pelanggan

Gambar Distribusi Segment Pelanggan menunjukkan jumlah pelanggan pada masing-masing segmen yang telah dibentuk berdasarkan hasil clustering. Berdasarkan grafik terlihat bahwa segmen Pelanggan Biasa memiliki jumlah pelanggan paling banyak dibandingkan segmen lainnya. Jumlah pelanggan pada segmen ini mencapai lebih dari tiga ribu pelanggan. Sementara itu, Pelanggan Pasif menempati posisi kedua dengan jumlah sekitar seribu pelanggan. Segmen ini menunjukkan adanya kelompok pelanggan yang sudah lama tidak melakukan transaksi dan berpotensi mengalami churn. Di sisi lain, jumlah Pelanggan Loyal merupakan yang paling sedikit dibandingkan segmen lainnya. Meskipun jumlah Pelanggan Loyal relatif kecil, kelompok ini memiliki kontribusi transaksi yang jauh lebih besar dibandingkan pelanggan lainnya. Kondisi ini menunjukkan bahwa sebagian besar pendapatan perusahaan berasal dari sejumlah kecil pelanggan yang memiliki tingkat loyalitas tinggi. Temuan tersebut sejalan dengan konsep Pareto Principle yang menyatakan bahwa sebagian besar hasil sering kali berasal dari sebagian kecil pelanggan bernilai tinggi.



Gambar 6 Distribusi Monetary per Segment

Gambar Distribusi Monetary per Segment memperlihatkan sebaran nilai pembelian pelanggan pada masing-masing segmen. Berdasarkan boxplot terlihat bahwa Pelanggan Loyal memiliki median dan rentang nilai Monetary yang jauh lebih tinggi dibandingkan segmen lainnya. Selain itu, terdapat beberapa outlier yang menunjukkan adanya pelanggan dengan nilai transaksi yang sangat besar. Pelanggan Biasa memiliki nilai Monetary yang berada pada tingkat menengah dengan variasi transaksi yang cukup beragam. Sebagian besar pelanggan pada segmen ini melakukan transaksi dalam jumlah yang relatif stabil namun belum mencapai tingkat pembelian pelanggan loyal. Sementara itu, Pelanggan Pasif memiliki nilai Monetary yang paling rendah. Hal ini menunjukkan bahwa pelanggan dalam segmen tersebut tidak hanya jarang melakukan transaksi tetapi juga memberikan kontribusi pendapatan yang relatif kecil. Hasil ini menunjukkan bahwa tingkat loyalitas pelanggan memiliki hubungan yang kuat dengan nilai transaksi yang dihasilkan.



Gambar 7 Segmentasi Pelanggan

(Sumbu X = Frequency, Sumbu Y = Monetary) Gambar Segmentasi Pelanggan menunjukkan hasil interpretasi cluster menjadi segmen bisnis yang lebih mudah dipahami, yaitu Pelanggan Loyal, Pelanggan Biasa, dan Pelanggan Pasif. Berdasarkan grafik terlihat bahwa Pelanggan Loyal (warna hijau) berada pada area dengan nilai Frequency dan Monetary yang paling tinggi. Hal ini menunjukkan bahwa pelanggan loyal melakukan transaksi secara rutin dan

memberikan kontribusi pendapatan terbesar bagi perusahaan. Pelanggan Biasa (warna oranye) merupakan kelompok yang paling dominan jumlahnya. Sebagian besar pelanggan pada segmen ini memiliki nilai Frequency dan Monetary pada tingkat menengah. Kelompok ini masih aktif melakukan transaksi dan memiliki potensi untuk ditingkatkan menjadi pelanggan loyal melalui strategi pemasaran yang tepat. Sementara itu, Pelanggan Pasif (warna biru) cenderung memiliki nilai Frequency yang rendah. Sebagian besar pelanggan dalam segmen ini melakukan transaksi dalam jumlah yang sedikit dan memiliki risiko lebih tinggi untuk berhenti bertransaksi. Oleh karena itu, perusahaan perlu menerapkan strategi reaktivasi pelanggan untuk mengurangi kemungkinan kehilangan pelanggan di masa mendatang.

D. PENUTUP

Kesimpulan

Penelitian ini berhasil menerapkan model RFM (Recency, Frequency, Monetary) bersama algoritma K-Means Clustering guna membagi pelanggan dalam set data Online Retail II. Dengan analisis metode Elbow Method, diketahui jumlah klaster optimal adalah tiga klaster yang mampu membedakan karakteristik pelanggan secara unik.

Proses pengklasteran mengungkapkan tiga segmen pelanggan utama: pelanggan loyal, pelanggan moderat, dan pelanggan kurang aktif. Pelanggan loyal menunjukkan nilai Frekuensi dan Moneter yang tinggi serta nilai Kebaruan yang rendah, menandakan frekuensi transaksi tinggi dan kontribusi pendapatan besar bagi bisnis. Sebaliknya, pelanggan kurang aktif memiliki nilai Kebaruan yang tinggi serta nilai Frekuensi dan Moneter yang rendah, berpotensi untuk beralih atau berhenti bertransaksi. Sementara itu, pelanggan moderat berada di tengah dengan aktivitas transaksi yang masih cukup stabil, berpeluang untuk dikembangkan menjadi pelanggan loyal.

Visualisasi hasil segmentasi melalui berbagai grafik seperti Elbow Method, Heatmap RFM, Distribusi Segment Pelanggan, Distribusi Monetary per Segment, dan Scatter Plot Segmentasi Pelanggan menunjukkan bahwa algoritma K-Means mampu memisahkan pelanggan berdasarkan karakteristik transaksi yang dimiliki. Selain itu, pengembangan dashboard customer intelligence memberikan kemudahan dalam memantau segmentasi pelanggan, mengidentifikasi pelanggan berisiko churn, menampilkan rekomendasi bundling produk, serta melakukan prediksi nilai transaksi. Hasil penelitian ini menunjukkan bahwa penerapan data mining menggunakan model RFM dan algoritma K-Means dapat menjadi solusi yang efektif dalam membantu perusahaan memahami perilaku pelanggan dan mendukung pengambilan keputusan bisnis yang lebih tepat sasaran. Dengan demikian, segmentasi pelanggan berbasis data dapat dimanfaatkan untuk meningkatkan efektivitas strategi pemasaran, retensi pelanggan, serta optimalisasi pendapatan perusahaan.

Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat dijadikan sebagai pengembangan pada penelitian selanjutnya. Penelitian ini menggunakan tiga atribut utama dalam model RFM, sehingga pada penelitian berikutnya dapat ditambahkan atribut lain yang relevan, seperti jenis produk yang sering dibeli, kategori pelanggan, maupun lokasi pelanggan untuk menghasilkan segmentasi yang lebih detail. Selain itu, metode clustering yang digunakan dapat dibandingkan dengan algoritma lain seperti Hierarchical Clustering, DBSCAN, atau Gaussian Mixture Model untuk mengetahui metode yang memberikan hasil segmentasi terbaik pada dataset yang sama.

Pengembangan dashboard customer intelligence juga dapat dilakukan dengan menambahkan fitur analisis secara real-time, integrasi dengan sistem Customer Relationship Management (CRM), serta implementasi notifikasi otomatis bagi pelanggan yang berisiko churn. Penelitian selanjutnya juga dapat memanfaatkan dataset yang lebih besar dan terbaru agar hasil segmentasi lebih representatif terhadap kondisi bisnis saat ini. Selain itu, evaluasi hasil clustering dapat dilengkapi dengan pengukuran kualitas cluster menggunakan metode seperti Silhouette Score atau Davies-Bouldin Index untuk memperoleh validasi yang lebih objektif. Dengan adanya pengembangan tersebut, diharapkan pemanfaatan data mining dalam segmentasi pelanggan dapat memberikan kontribusi yang lebih optimal dalam mendukung strategi bisnis berbasis data pada perusahaan retail maupun sektor bisnis lainnya.

E. DAFTAR PUSTAKA

- Tivano, D., Ruslianto, I., & Prawira, D. (2023). Sistem Segmentasi Pelanggan Potensial Menggunakan Metode K-Means dan Analisis RFM.
- Fadhillah, M. F., Suyoso, A. L. A., & Puspitasari, I. (2024). Segmentasi Pelanggan dengan Algoritma Clustering Berdasarkan Atribut Recency, Frequency dan Monetary (RFM).
- Passolving, I. Y., Ali, P. R., Kartawidjaja, M. A., & Sukwadi, R. (2024). Segmentasi Pelanggan Menggunakan K-Means Clustering: Menganalisis Metrik RFM untuk Strategi Pemasaran.
- Lukas, I., Okmayura, F., Irani, A. T., dkk. (2025). Pemodelan RFM & K-Means Clustering untuk Segmentasi Pelanggan dalam Penjualan Online.
- Nurhidayat, A. F., Kusuma, M. J. L., & Septiawan, R. (2025).
- Penerapan Teknik Clustering Menggunakan Algoritma KMeans Berbasis RFM untuk Segmentasi Pelanggan pada Dataset Online Retail II.

Dharma, N. C., & Fenriana, I. (2024). Perancangan Aplikasi Segmentasi Pelanggan Menggunakan KMeans Dengan Model RFM.