

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi digital telah secara signifikan mengubah cara manusia berinteraksi, bertransaksi, dan menjalankan aktivitas ekonomi. Salah satu dampak paling mencolok dari perkembangan teknologi ini adalah munculnya *e-commerce* sebagai model bisnis yang mendominasi perdagangan. Berdasarkan data Statista, nilai penjualan *e-commerce* global diperkirakan mencapai US\$6,42 triliun pada tahun 2025 [1], meningkat signifikan dibandingkan tahun 2024 sebesar US\$3,6 triliun [2]. Tren serupa juga terjadi di Indonesia, di mana nilai transaksi *e-commerce* mencapai Rp1.775 triliun pada tahun 2025 [3] dan meningkat sebesar 264,47% dari tahun sebelumnya yakni sebesar Rp487 triliun [4]. Pertumbuhan pesat ini utamanya didorong oleh kombinasi beberapa faktor, seperti perubahan gaya hidup modern, meluasnya adopsi internet, dan inovasi dalam sistem pembayaran digital [5].

Pertumbuhan *e-commerce* tidak hanya tercermin dari peningkatan nilai transaksi, tetapi juga dari perubahan peran dan fungsinya dalam perekonomian digital. *E-commerce* yang sebelumnya berfungsi sebagai elemen pelengkap kini bertransformasi menjadi elemen sentral [6]. Transformasi ini menjadikan *e-commerce* berperan tidak hanya sebagai saluran distribusi barang dan jasa, tetapi juga berperan sebagai penggerak inovasi, pencipta lapangan kerja, dan pendorong pertumbuhan ekonomi digital di berbagai negara. Namun, dibalik pertumbuhan tersebut, muncul tantangan baru dalam pengelolaan konsumen, khususnya terkait perbedaan perilaku dan pola konsumsi konsumen di tengah persaingan yang semakin ketat. Perubahan tersebut menunjukkan bahwa setiap konsumen memiliki karakteristik, kebutuhan, dan nilai yang berbeda, sehingga pendekatan pemasaran yang bersifat umum tidak lagi efektif [7]. Oleh karena itu, diperlukan strategi pemasaran berbasis segmentasi konsumen sebagai upaya untuk mengelompokkan konsumen berdasarkan karakteristik dan perilaku yang relatif homogen, sehingga strategi yang diterapkan dapat lebih tepat sasaran dan efektif dalam menjawab tantangan tersebut.

Sejalan dengan hal itu, analisis perilaku konsumen berbasis data menjadi penting untuk memahami dinamika pasar digital yang semakin kompleks. Salah satu metode yang dapat digunakan adalah model *Recency, Frequency*, dan *Monetary* (RFM). Model ini memungkinkan perusahaan mengelompokkan konsumen berdasarkan seberapa baru konsumen melakukan transaksi (*Recency*), seberapa sering mereka bertransaksi (*Frequency*), serta berapa besar nilai

transaksinya (*Monetary*) [8]. Melalui pengelompokan ini, analisis tersebut menjadi sangat bermanfaat untuk mengidentifikasi konsumen potensial, konsumen setia, maupun konsumen yang mulai tidak aktif.

Namun demikian, model RFM memiliki keterbatasan karena hanya memberikan gambaran statis dan tidak mampu menangkap dinamika perilaku konsumen yang berubah seiring waktu. Banyak penelitian terkini yang mengintegrasikan analisis RFM dengan algoritma *clustering* berbasis machine learning untuk mengatasi keterbatasan tersebut. Misalnya, kombinasi RFM dan algoritma *K-Means* mampu menghasilkan segmentasi konsumen yang lebih terarah berdasarkan kecenderungan belanja dan sensitivitas harga [9]. Fakta lain menunjukkan bahwa integrasi RFM dengan algoritma seperti DBSCAN dan *Gaussian Mixture Model* dapat meningkatkan efektivitas strategi pemasaran berbasis data serta memprediksi potensi perilaku pembelian konsumen di masa depan dengan tingkat akurasi lebih tinggi [10]. Sementara itu, penelitian lain menunjukkan bahwa integrasi RFM dengan algoritma *Mean-Shift* efektif dalam mengidentifikasi jumlah segmen konsumen yang paling alami berdasarkan kepadatan data, sehingga mampu mengungkap kelompok-kelompok tersembunyi tanpa asumsi awal [11]. Temuan-temuan tersebut membuktikan bahwa pendekatan integratif ini dapat menjadi alat yang kuat dalam pengambilan keputusan strategis.

Meskipun berbagai penelitian telah menunjukkan keberhasilan integrasi model RFM dengan algoritma *clustering* tertentu, sebagian besar masih berfokus pada penerapan satu algoritma secara terpisah. Padahal, setiap algoritma memiliki karakteristik dan keunggulan yang berbeda dalam mengelompokkan data konsumen. *K-Means* efektif untuk data dengan bentuk cluster yang teratur [9], GMM mampu mengenali pola dengan distribusi elips [10], DBSCAN unggul dalam mendeteksi cluster berdasarkan kepadatan dan mengidentifikasi outlier [10], sedangkan *Mean-Shift* dapat menemukan jumlah cluster secara otomatis tanpa asumsi awal [11]. Perbedaan karakteristik tersebut menunjukkan bahwa tidak ada satu algoritma yang bersifat universal untuk semua jenis data konsumen. Akibatnya, Ketergantungan pada satu algoritma tunggal ini berisiko menghasilkan segmentasi yang tidak stabil dan kurang representatif terhadap struktur data yang sebenarnya [12].

Dari hasil penelusuran literatur, penelitian yang secara sistematis membandingkan kinerja dari beberapa model seperti *K-Means*, GMM, DBSCAN dan Mean Shift pada data RFM *e-commerce* masih jarang dilakukan. Kesenjangan inilah yang mendasari diusulkannya penelitian berjudul “Eksplorasi Pola Segmentasi Konsumen *E-commerce* Menggunakan Integrasi Analisis RFM dan Algoritma *Clustering*”. Penelitian ini bertujuan untuk mengevaluasi dan membandingkan beberapa pendekatan integratif tersebut untuk

mengidentifikasi algoritma mana yang menghasilkan segmentasi konsumen paling valid dan optimal. Hasilnya diharapkan dapat memberikan kontribusi praktis bagi pelaku *e-commerce* dalam merancang strategi pemasaran yang lebih efektif, efisien, dan personal berbasis data perilaku konsumen .

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana penerapan model RFM sebagai dasar analisis perilaku *konsumen e-commerce* sebelum dilakukan pengelompokan menggunakan algoritma *clustering*?
2. Bagaimana penerapan dan perbandingan berbagai algoritma *clustering* seperti *K-Means*, DBSCAN, GMM, dan *Mean-Shift* dalam membentuk segmentasi konsumen *e-commerce* berdasarkan karakteristik data konsumen ?
3. Algoritma *clustering* manakah yang memberikan hasil segmentasi terbaik berdasarkan metrik evaluasi kuantitatif (*Silhouette Score*, DBI, dan CHI)?.
4. Bagaimana hasil segmentasi konsumen tersebut dapat digunakan untuk menyusun rekomendasi strategi bisnis berbasis data yang relevan bagi pelaku *e-commerce*?

1.3 Tujuan

Berdasarkan rumusan masalah yang telah diuraikan, tujuan dari penelitian ini adalah sebagai berikut:

1. Menerapkan model RFM sebagai dasar analisis untuk menggambarkan dan mengukur perilaku konsumen *e-commerce* secara kuantitatif sebelum dilakukan proses pengelompokan..
2. Menerapkan dan membandingkan berbagai algoritma *clustering*, yaitu *K-Means*, DBSCAN, *Gaussian Mixture Model* (GMM), dan *Mean-Shift*, dalam membentuk segmentasi konsumen *e-commerce* berdasarkan karakteristik data konsumen
3. Mengevaluasi kualitas hasil segmentasi berdasarkan metrik *Silhouette Score*, *Davies–Bouldin Index*, dan *Calinski–Harabasz Index* guna menentukan algoritma paling efektif
4. Menyusun rekomendasi strategi berbasis data konsumen sebagai dasar pengambilan keputusan bisnis yang lebih tepat dan efisien di sektor *e-commerce*

1.4 Manfaat

Berdasarkan rumusan masalah dan tujuan penelitian yang telah disampaikan, manfaat dari penelitian ini adalah sebagai berikut:

1. Menghasilkan segmentasi konsumen *e-commerce* berbasis analisis RFM yang mampu menggambarkan perilaku konsumen secara objektif dan terukur sebagai dasar analisis lebih lanjut.
2. Memberikan perbandingan kinerja berbagai algoritma *clustering*, yaitu *K-Means*, DBSCAN, *Gaussian Mixture Model* (GMM), dan *Mean-Shift*, dalam membentuk segmentasi konsumen *e-commerce* berdasarkan karakteristik data dan hasil evaluasi *clustering*.
3. Memberikan rekomendasi strategi bisnis berbasis data untuk meningkatkan efektivitas pemasaran, retensi, dan loyalitas konsumen.
4. Mendukung pengambilan keputusan bisnis berbasis data (*data-driven decision making*) pada sektor *e-commerce* melalui hasil analisis dan segmentasi konsumen yang akurat.

1.5 Ruang Lingkup

Berdasarkan fokus penelitian yang telah dirumuskan, ruang lingkup penelitian ini meliputi:

1. Dataset transaksi konsumen *e-commerce* yang diambil dari platform Kaggle <https://www.kaggle.com/datasets/shriyashjagtap/e-commerce-customer-for-behavior-analysis>
2. Metode yang digunakan yakni *Knowledge Discovery in Databases* (KDD) yang mencakup tahapan data *Selection*, *Preprocessing*, transformation, data mining, dan evaluation. Pada tahap data mining, diterapkan analisis komparatif untuk menguji dan membandingkan kinerja dari beberapa algoritma *clustering*, yaitu *K-Means*, DBSCAN, *Gaussian Mixture Model* (GMM), dan *Mean-Shift*.
3. Variabel yang dianalisis adalah *Recency*, *Frequency*, dan *Monetary* (RFM) yang dihasilkan dari data transaksi.
4. Analisis difokuskan pada perbandingan hasil segmentasi dan interpretasi karakteristik konsumen dari tiap kelompok.
5. Tidak mencakup analisis demografi, psikografis, atau implementasi teknis strategi pemasaran pada level operasional.