

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi digital mendorong peningkatan volume data secara signifikan, khususnya dalam industri ritel dan *e-commerce*. Data transaksi konsumen yang terus bertambah setiap harinya menyimpan informasi berharga mengenai pola perilaku konsumen. Namun, data mentah tidak akan memberikan nilai strategis apabila tidak diolah [1]. *Data mining* menawarkan sebuah solusi untuk mengekstraksi wawasan dan pola yang tidak terlihat secara langsung dari kumpulan data, yang pada gilirannya berfungsi untuk menunjang proses pengambilan keputusan yang berbasis data [2].

Salah satu teknik *data mining* yang banyak digunakan dalam analisis konsumen adalah *clustering*, yaitu pengelompokan data berdasarkan kesamaan karakteristik. *Clustering* bermanfaat dalam segmentasi konsumen karena memungkinkan perusahaan mengidentifikasi kelompok konsumen dengan perilaku yang serupa, sehingga strategi pemasaran dapat lebih tepat sasaran [3]. Dari berbagai metode *clustering*, K-Means merupakan algoritma yang paling populer karena kesederhanaannya, efisiensi komputasi, serta kemampuannya dalam menangani data berskala besar [4].

Dalam penelitian ini, penentuan jumlah cluster dilakukan dengan memanfaatkan beberapa metrik evaluasi internal, yaitu *Elbow Method*, *Silhouette Score*, *Davies-Bouldin Index (DBI)*, dan *Calinski-Harabasz Score (CHS)* [5]. Keempat metrik tersebut digunakan secara bersamaan untuk memperoleh jumlah *cluster* yang optimal. Selain itu, penerapan RFM (*Recency, Frequency, Monetary*) menjadi salah satu pendekatan yang efektif dalam segmentasi konsumen. Hal pertama yang perlu dilakukan adalah mengubah data perusahaan menjadi *recency*, *frequency*, dan *monetary* (RFM). RFM memungkinkan pengelompokan konsumen berdasarkan seberapa baru mereka melakukan transaksi (*recency*), seberapa sering bertransaksi (*frequency*), dan seberapa besar nilai transaksi yang dilakukan (*monetary*) [6]. Dengan menggunakan RFM sebagai basis data input, algoritma K-Means dapat menghasilkan profil konsumen yang lebih representatif dan relevan secara bisnis.

Untuk memastikan bahwa *cluster* yang terbentuk memiliki perbedaan signifikan secara statistik, evaluasi tambahan dilakukan menggunakan *Analysis of Variance* (ANOVA) atau *one way* ANOVA. Metode ini membandingkan variasi antar *cluster* sehingga dapat divalidasi bahwa perbedaan yang terbentuk bukan sekadar akibat variasi acak, melainkan benar-benar mencerminkan segmentasi konsumen yang bermakna [7].

1.2 Perumusan Masalah Penelitian

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, perumusan masalah dalam penelitian ini adalah:

1. Bagaimana penerapan algoritma *K-Means* dapat digunakan untuk melakukan segmentasi konsumen sehingga menghasilkan profil *cluster* yang merepresentasikan perbedaan perilaku konsumen dan dapat dijadikan dasar strategi bisnis?
2. Bagaimana kombinasi evaluasi multi-metrik (*Elbow Method*, *Silhouette Score*, *Davies-Bouldin Index*, dan *Calinski-Harabasz Score*) serta validasi menggunakan *Analysis of Variance* (ANOVA) dapat digunakan untuk menentukan jumlah *cluster* yang optimal sekaligus memastikan kualitas segmentasi yang dihasilkan?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah untuk:

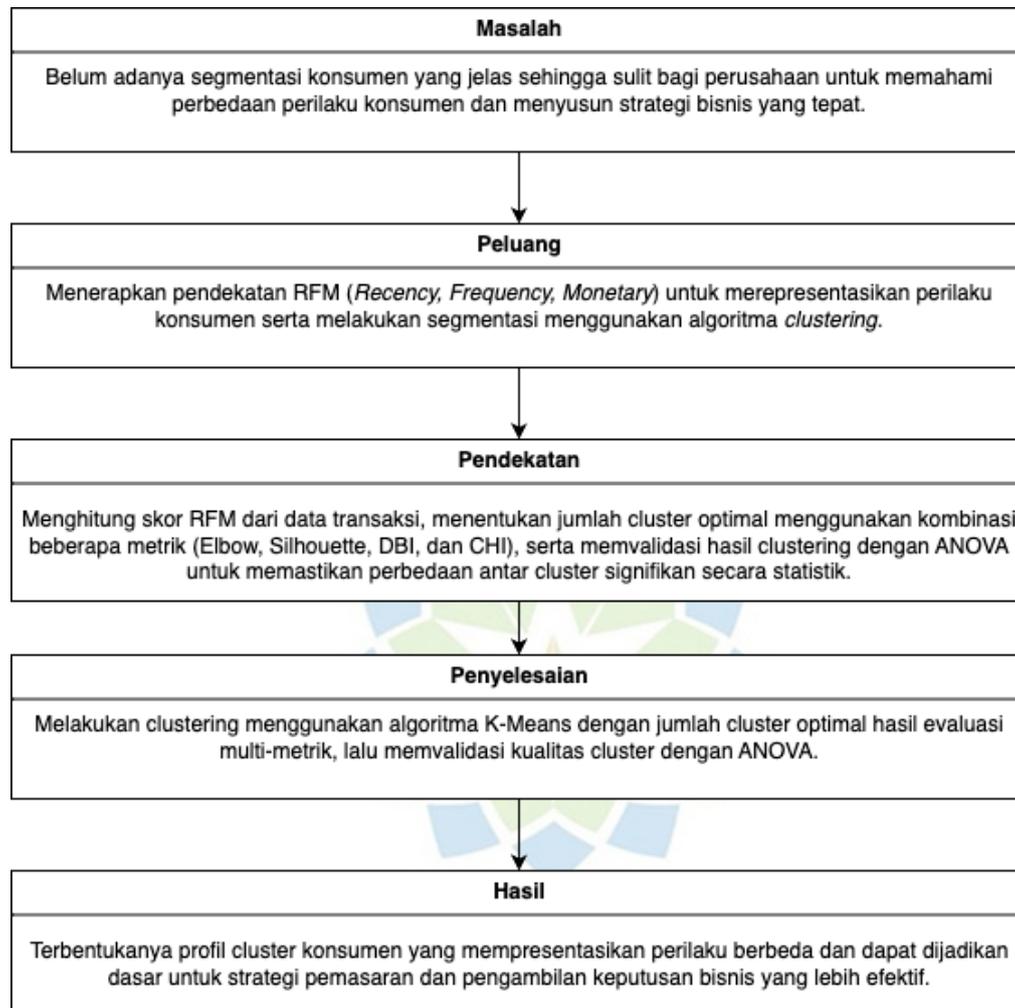
1. Menerapkan algoritma *K-Means* untuk melakukan segmentasi konsumen sehingga diperoleh profil *cluster* yang merepresentasikan perbedaan perilaku konsumen dan dapat dijadikan dasar dalam perumusan strategi bisnis.
2. Mengevaluasi hasil segmentasi menggunakan kombinasi metrik *Elbow Method*, *Silhouette Score*, *Davies-Bouldin Index*, dan *Calinski-Harabasz Score*, serta melakukan validasi dengan *Analysis of Variance* (ANOVA) guna menentukan jumlah *cluster* yang optimal dan memastikan kualitas segmentasi yang dihasilkan.

1.4 Batasan Masalah Penelitian

Agar tema pada penelitian ini tidak keluar dari jalurnya, maka diperlukan pembatasan masalah yang diantara adalah sebagai berikut:

- a. Penelitian ini difokuskan pada CV Noorani Jaya, sebuah perusahaan percetakan, dengan menggunakan data transaksi pembelian konsumen selama 2 tahun. Data yang dianalisis berbentuk tabular, mencakup waktu pembelian, kategori produk, jumlah pembelian, dan nama konsumen. Aspek lain seperti keuangan, operasional, pemasok, preferensi pribadi, demografi, atau perilaku digital konsumen tidak disertakan karena tidak tersedia dalam sistem.
- b. Segmentasi konsumen dilakukan dengan menggunakan pendekatan RFM (Recency, Frequency, Monetary) sebagai dasar perhitungan variabel input dalam algoritma K-Means. Analisis hanya difokuskan pada pengelompokan konsumen berdasarkan RFM tanpa melibatkan variabel non-transaksional.
- c. Penelitian ini hanya mencakup tahap analisis data hingga validasi hasil clustering menggunakan metrik evaluasi internal dan ANOVA, tanpa mencakup deployment, pengembangan sistem rekomendasi, atau integrasi ke dalam sistem informasi perusahaan.
- d. Seluruh hasil analisis dan rekomendasi dalam penelitian ini bersifat dukungan keputusan internal dan digunakan sebagai masukan dalam strategi pemasaran digital CV Noorani Jaya. Rekomendasi yang diberikan tidak bersifat wajib maupun menggantikan strategi pemasaran yang telah ada.

1.5 Kerangka Pemikiran Penelitian



Gambar 1.1 Kerangka Pemikiran Penelitian

1.6 Sistematika Penulisan

Sistematika penulisan penelitian ini disusun untuk memberikan gambaran umum tentang proses penulisan yang dilakukan untuk menyelesaikan penelitian ini. Sistematika penulisan penelitian ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini terdiri dari latar belakang, perumusan masalah penelitian, tujuan penelitian, batasan masalah penelitian, kerangka pemikiran penelitian, metodologi penelitian, dan sistematika penulisan.

BAB II STUDI PUSTAKA

Bab ini terdiri dari pembahasan konsep dasar dan teori-teori yang berkaitan dengan topik masalah yang diambil dan hal-hal yang berguna dalam proses analisis permasalahan.

BAB III METODOLOGI PENELITIAN

Bab ini terdiri dari pembahasan tentang alur metodologi penelitian dari tahap awal sampai akhir yang di dalamnya terdiri dari pembahasan analisa produk dan perencanaan eksekusi aplikasi.

BAB IV HASIL DAN PEMBAHASAN

Bab ini terdiri dari dua hal utama, pertama pemaparan tentang temuan atau hasil penelitian berdasarkan tahapan penelitian yang dilakukan. Pemaparan hasil penelitian disesuaikan dengan urutan masalah penelitian. Kedua pembahasan hasil atau temuan penelitian untuk menjawab rumusan penelitian.

BAB V PENUTUP

Bab ini terdiri dari pembahasan kesimpulan penelitian, kritik, dan saran yang penulis dapatkan selama menyelesaikan penelitian tugas akhir ini. Intisari dari bab ini dapat digunakan sebagai pertimbangan untuk pengembangan penulisan selanjutnya.