

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Sektor manufaktur merupakan salah satu penggerak utama perekonomian Indonesia dengan kontribusi terhadap Produk Domestik Bruto (PDB) yang tertinggi dibandingkan dengan sektor-sektor lainnya. Namun, dalam kurun satu dekade terakhir, kontribusi tersebut justru mengalami tren penurunan, dari 23,69% pada tahun 2013 menjadi 18,64% pada tahun 2023 [1]. Berbeda dengan tren makro tersebut, industri fashion sebagai bagian dari sektor manufaktur mampu mempertahankan eksistensinya, terutama pada skala Industri Mikro dan Kecil (IMK). Skala industri ini mayoritas digerakkan oleh pelaku Usaha Mikro, Kecil, dan Menengah (UMKM) yang memiliki jangkauan pasar yang luas di seluruh Indonesia. Berdasarkan Klasifikasi Baku Lapangan Usaha Indonesia (KBLI) 14, industri pakaian jadi menempati posisi kedua terbanyak dari seluruh kelompok KBLI di Provinsi Banten, dengan total 15.453 unit usaha [2]. Besarnya jumlah pelaku usaha ini sejalan dengan tingginya daya beli masyarakat Banten. Hal ini tecermin dari pengeluaran konsumsi rumah tangga yang mendominasi struktur Produk Domestik Regional Bruto (PDRB) Provinsi Banten dari sisi pengeluaran, yaitu mencapai 52,47% pada tahun 2023 [3]. Kondisi tersebut menciptakan peluang pasar yang luas bagi produk fashion, sekaligus menuntut pelaku usaha untuk memiliki kemampuan dalam menjangkau segmen konsumen secara tepat sasaran.

Meskipun, peluang pasar yang luas tersebut belum sepenuhnya dapat dioptimalkan, khususnya oleh pelaku UMKM fashion yang baru berada pada tahap merintis usaha. Pada fase awal ini, pelaku usaha umumnya belum memiliki basis data pelanggan yang memadai sebagai landasan pengambilan keputusan. Akibatnya, penentuan target konsumen hanya didasarkan pada asumsi tanpa didukung oleh riset perilaku konsumen. Kondisi ini mengakibatkan produk yang ditawarkan tidak relevan dengan kebutuhan pasar. Permasalahan ini sejalan dengan temuan Evermos yang dipublikasikan melalui *World Economic Forum*, yang mencatat bahwa lebih dari 99% pelaku usaha tahap awal (*newcomer*) di Indonesia mengalami stagnasi akibat ketidakmampuan mengidentifikasi target pasar secara presisi serta menentukan jenis produk yang benar-benar dibutuhkan [4]. Persoalan ini semakin kompleks apabila dikaitkan dengan karakteristik geografis Provinsi

Banten yang mencakup delapan kabupaten/kota dengan tingkat daya beli yang bervariasi. Data Badan Pusat Statistik (BPS) menunjukkan bahwa Kota Tangerang dan Kota Tangerang Selatan memiliki rata-rata pengeluaran per kapita tertinggi, sedangkan Kabupaten Pandeglang dan Kabupaten Lebak berada pada posisi terendah, sehingga mencerminkan adanya perbedaan pola konsumsi yang cukup signifikan antarwilayah [5].

Dalam beberapa tahun terakhir, penelitian mengenai segmentasi konsumen menggunakan metode *K-Means clustering* cukup banyak dilakukan dalam literatur nasional, baik pada konteks UMKM lokal maupun pada *dataset* acuan yang lebih umum. Kristanti et al. [6] mengelompokkan 1.140 data pelanggan pada sebuah toko ritel di Kabupaten Jombang dengan memadukan atribut usia dan pendapatan bersama komponen RFM, sebelum akhirnya menetapkan empat segmen pelanggan dengan karakteristik usia dan daya beli yang berlainan. Pendekatan serupa diterapkan oleh Ardi et al. [7] pada Fashion Viral Solo, yang menghasilkan dua kelompok pelanggan dengan komposisi anggota yang jauh tidak seimbang (343 berbanding 8 pelanggan). Pola yang sama juga terlihat pada penelitian Perdana et al. [8] terhadap 19.492 pengguna aplikasi Alfagift berdasarkan histori transaksi Juni 2021, yang menggunakan pendekatan *CRISP-DM* dan metode *Elbow* untuk memperoleh tiga *cluster* pelanggan (7.219, 6.902, dan 5.371 pelanggan), serta Ahsina et al. [9] yang mengelompokkan 1.000 data nasabah bank berdasarkan riwayat pengambilan kredit menjadi empat *cluster* dengan metode *Elbow*. Kelemahan utama dari keempat pendekatan ini terletak pada sumber datanya yang sepenuhnya berasal dari riwayat transaksi pelanggan yang sudah pernah berbelanja atau bertransaksi, sehingga model yang terbentuk hanya mampu menggambarkan pola yang telah terjadi dan tidak dapat diadopsi oleh pelaku usaha yang baru merintis bisnis dan belum memiliki catatan transaksi apa pun, termasuk UMKM fashion pemula yang menjadi sasaran penelitian ini.

Pada konteks yang sedikit berbeda, Zaenal et al. [10] melakukan segmentasi terhadap pelanggan toko MDA Collection dengan menggabungkan algoritma *K-Means* dan analisis RFM, yang menghasilkan empat kelompok pelanggan berdasarkan tingkat loyalitasnya. Segmentasi berbasis loyalitas ini cukup membantu pihak toko dalam merancang program retensi pelanggan, namun seperti penelitian sebelumnya, seluruh proses pengelompokan tetap bergantung pada data transaksi pelanggan yang telah tercatat sebelumnya, sehingga model ini juga tidak dapat dimanfaatkan oleh pelaku usaha yang belum memiliki riwayat penjualan maupun basis pelanggan tetap.

Sejumlah penelitian *K-Means* lain justru memanfaatkan *dataset* publik yang banyak digunakan untuk kepentingan pembelajaran maupun eksperimen metodologis, salah satunya *Mall Customers* yang tersedia secara terbuka di platform Kaggle. Muhammad et al. [11] menerapkan *K-Means* pada *dataset* tersebut dengan memanfaatkan variabel pendapatan tahunan dan skor pengeluaran pelanggan, memperoleh lima kelompok pelanggan melalui evaluasi *Davies–Bouldin Index*. Pendekatan yang identik juga diterapkan oleh Siagian et al. [12] pada *dataset* yang sama, yang turut menghasilkan lima *cluster* pelanggan (81, 39, 22, 35, dan 23 pelanggan). Penggunaan *dataset* sintetis yang bersifat generik dan tidak terikat pada konteks usaha maupun wilayah geografis tertentu pada kedua penelitian ini membuat hasilnya kurang dapat digeneralisasi pada permasalahan nyata yang dihadapi UMKM fashion di suatu daerah spesifik. Fenomena serupa juga terdapat pada penelitian yang dilakukan oleh Alkautsar et al. [13] yang menerapkan *K-Means* untuk mengelompokkan mahasiswa berdasarkan potensi waktu kelulusan menggunakan data IPK dan SKS. Menariknya, hasil evaluasi pada penelitian tersebut menunjukkan inkonsistensi antarmetrik: *Davies–Bouldin Index* (0,67) dan *Adjusted Rand Index* (0,17) menyatakan hasil klaster tidak memuaskan, sedangkan *Calinski–Harabasz Index* (338,28) menyatakan sebaliknya. Temuan ini menegaskan bahwa evaluasi *cluster* menggunakan satu metrik saja berisiko memberikan kesimpulan yang bias, sehingga memperkuat perlunya pendekatan multi-metrik seperti yang diterapkan dalam penelitian ini.

Pemilihan algoritma *K-Means* dibandingkan metode *clustering* lain juga perlu diamati. Ripai et al. [14] membandingkan algoritma *K-Means*, *Fuzzy C-Means*, dan *Hierarchical Clustering* dalam mengelompokkan tingkat pemahaman siswa berdasarkan nilai akademik, dan menemukan bahwa *K-Means* memberikan hasil klasterisasi yang lebih baik dibandingkan kedua metode pembanding, dengan nilai *Silhouette Index* 0,350, *Davies–Bouldin Index* 0,822, dan *Calinski–Harabasz Index* 1157,8. Sementara itu, Hartono et al. [15] membandingkan proses klasterisasi *K-Means* dan *Agglomerative Hierarchical Clustering* (AHC) pada tiga ukuran data berbeda (10, 30, dan 60 titik data), dan menemukan bahwa kedua metode menghasilkan klaster yang identik pada data berjumlah kecil, namun mulai menunjukkan perbedaan hasil seiring bertambahnya jumlah data. Kedua temuan ini menjadi pertimbangan tambahan bagi penelitian ini dalam memilih *K-Means*, mengingat karakteristik data survei primer yang berjumlah relatif besar (1.189 responden) dan bersifat numerik multidimensi, sehingga kestabilan hasil klaster pada skala data yang lebih besar menjadi pertimbangan penting.

Kedelapan penelitian tersebut, meskipun berbeda dalam hal studi kasus dan sumber data, memiliki kelemahan yang dapat dikelompokkan ke dalam dua persoalan utama. Pertama, penelitian yang menggunakan data transaksi riil seperti Kristanti et al. [6], Ardi et al. [7], Zaenal et al. [10], Perdana et al. [8], dan Ahsina et al. [9] seluruhnya bertumpu pada riwayat transaksi pelanggan yang sudah tercatat sebelumnya, sehingga tidak dapat diterapkan oleh pelaku UMKM fashion pemula yang sama sekali belum memiliki data pembelian. Kedua, penelitian yang berusaha keluar dari ketergantungan pada data transaksi seperti Muhammad et al. [11], Siagian et al. [12], dan Alkautsar et al. [13] justru beralih ke *dataset* publik atau domain lain yang generik dan tidak merepresentasikan konteks usaha, perilaku konsumen, maupun kondisi geografis wilayah tertentu. Ketiga, tidak satu pun dari kedelapan penelitian tersebut mengintegrasikan hasil segmentasi konsumen dengan data daya beli wilayah secara eksplisit. Akibatnya, segmen-segmen yang terbentuk hanya menggambarkan karakteristik perilaku konsumen secara abstrak tanpa dapat menunjukkan di wilayah mana segmen tersebut paling banyak ditemukan. Hingga saat ini, penelitian yang secara khusus menggunakan data primer berbasis survei untuk membangun segmentasi konsumen fashion, kemudian mengaitkan hasilnya dengan indikator daya beli wilayah dari BPS untuk menghasilkan rekomendasi target pasar yang spesifik secara geografis, belum ditemukan dalam literatur yang membahas konteks Provinsi Banten maupun UMKM fashion pemula secara umum.

Berdasarkan kondisi tersebut, penelitian ini menggunakan algoritma K-Means Clustering sebagai metode utama untuk melakukan segmentasi konsumen, yang dipilih karena mampu mengelompokkan data multidimensi secara efisien tanpa memerlukan label kelas sebelumnya (*unsupervised learning*), sehingga cocok untuk menemukan pola tersembunyi dalam data perilaku konsumen yang bersumber dari survei primer melalui kuesioner terstruktur. Hasil segmentasi kemudian akan dikombinasikan dengan data pengeluaran per kapita dari Badan Pusat Statistik pada tingkat kabupaten dan kota di Provinsi Banten, sehingga setiap segmen konsumen yang terbentuk dapat dikaitkan dengan wilayah yang memiliki tingkat daya beli yang sesuai. Untuk menentukan jumlah cluster yang paling optimal, penelitian ini menggunakan kombinasi metode Elbow dan Silhouette, sedangkan kualitas hasil clustering akan dievaluasi menggunakan tiga metrik validasi, yaitu Silhouette Score, Davies-Bouldin Index (DBI), dan Calinski-Harabasz Index (CH). Seluruh proses analisis serta hasil rekomendasi segmentasi yang dihasilkan akan dikemas dalam model yang dirancang untuk memberikan panduan pemetaan target pasar berdasarkan karakteristik konsumen dan daya beli wilayah, yang selanjutnya dapat

dimanfaatkan secara langsung oleh pelaku UMKM fashion pemula di Provinsi Banten.

Setidaknya, terdapat tiga permasalahan utama yang menjadi fokus dalam penelitian ini. Pertama, belum tersedianya model segmentasi target pasar berbasis survei primer yang dapat dimanfaatkan oleh pelaku UMKM fashion pemula di Provinsi Banten yang belum memiliki data pelanggan. Kedua, belum diketahui secara pasti jumlah cluster optimal serta profil karakteristik dari masing-masing segmen konsumen fashion di wilayah tersebut. Ketiga, belum tersedia model rekomendasi yang mengintegrasikan profil segmen konsumen dengan karakteristik ekonomi wilayah berbasis data survei pada delapan kabupaten/kota di Provinsi Banten, dengan data pengeluaran per kapita BPS sebagai pembanding untuk menghasilkan rekomendasi target pasar yang spesifik secara geografis bagi pelaku UMKM *fashion* pemula.

1.2 Rumusan Masalah

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

1. Bagaimana penerapan algoritma *K-Means Clustering* dalam melakukan segmentasi target pasar produk fashion untuk UMKM pemula di Provinsi Banten berdasarkan data survei primer?
2. Berapa jumlah *cluster* optimal dan bagaimana profil karakteristik setiap segmen konsumen fashion di Provinsi Banten yang dihasilkan oleh algoritma *K-Means Clustering*?
3. Bagaimana mengintegrasikan hasil segmentasi konsumen dengan karakteristik ekonomi wilayah berbasis data survei pada delapan kabupaten/kota di Provinsi Banten, dengan data pengeluaran per kapita BPS sebagai pembanding, untuk menghasilkan rekomendasi target pasar yang spesifik secara geografis bagi UMKM *fashion* pemula?

1.3 Batasan Permasalahan

Berdasarkan ruang lingkup penelitian, maka batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Penelitian berfokus pada segmentasi konsumen produk fashion yang berdomisili di delapan kabupaten/kota di Provinsi Banten.
2. Data yang digunakan dalam penelitian ini bersumber dari survei primer yang dikumpulkan melalui kuesioner daring.
3. Algoritma yang digunakan adalah *K-Means Clustering*. Penentuan jumlah *cluster* optimal menggunakan *Elbow Method* dan *Silhouette Method*, serta dievaluasi menggunakan *Silhouette Score*, *Davies-Bouldin Index* (DBI), dan *Calinski-Harabasz Index* (CH).

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah dalam penelitian ini, tujuan yang hendak dicapai adalah sebagai berikut:

1. Menerapkan algoritma *K-Means Clustering* untuk membangun model segmentasi target pasar produk fashion bagi UMKM pemula di Provinsi Banten berdasarkan data survei primer.
2. Menentukan jumlah *cluster* optimal serta menganalisis profil karakteristik setiap segmen konsumen fashion di Provinsi Banten yang dihasilkan oleh algoritma *K-Means Clustering*.
3. Mengintegrasikan hasil segmentasi konsumen dengan karakteristik ekonomi wilayah berbasis data survei kabupaten/kota di Provinsi Banten, dengan data pengeluaran per kapita BPS sebagai pembanding, untuk menghasilkan rekomendasi target pasar yang spesifik secara geografis bagi UMKM fashion pemula.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat bagi berbagai pihak, baik secara teoritis maupun praktis, sebagai berikut :

A Manfaat Teoritis

1. Memberikan kontribusi konseptual mengenai penerapan algoritma *K-Means Clustering* yang dipadukan dengan pengujian stabilitas *cluster*, sehingga

dapat menjadi acuan bagi penentuan jumlah cluster yang lebih andal dibandingkan hanya mengandalkan Elbow Method atau Silhouette Method, khususnya pada segmentasi pasar UMKM fashion pemula yang belum memiliki riwayat transaksi.

2. Memperkaya literatur mengenai integrasi segmentasi konsumen dengan karakteristik ekonomi wilayah berbasis data survei sebagai dasar rekomendasi target pasar geografis, dan data BPS sebagai pembanding untuk UMKM pemula.
3. Menjadi referensi bagi penelitian selanjutnya yang mengkaji segmentasi pasar fashion di Provinsi Banten maupun daerah lain dengan karakteristik yang serupa.

B Manfaat Praktis

1. Bagi UMKM fashion pemula: memberikan referensi berbasis data mengenai profil segmen konsumen dan wilayah target pasar yang potensial di Provinsi Banten.
2. Bagi konsumen fashion di Banten: mendorong UMKM untuk menyesuaikan produk dengan kebutuhan konsumen yang sesungguhnya di tiap wilayah.
3. Bagi pemerintah daerah dan dinas terkait: hasil penelitian dapat dijadikan masukan dalam program pembinaan UMKM fashion berbasis data di Provinsi Banten.
4. Bagi peneliti, penelitian ini menjadi sarana penerapan ilmu informatika, mulai dari praproses data, pemodelan algoritma *K-Means Clustering*, hingga uji validitas dan stabilitas model, sehingga dapat meningkatkan kemampuan peneliti dalam menyelesaikan permasalahan nyata di lingkup UMKM lokal.

1.6 Sistematika Penulisan

Sistematika penulisan laporan penelitian ini adalah sebagai berikut:

1. Bab 1 PENDAHULUAN

Bab ini berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan.

2. Bab 2 LANDASAN TEORI

Bab ini berisi landasan teori yang mendukung penelitian, meliputi konsep segmentasi pasar, algoritma *K-Means Clustering*, metode penentuan jumlah *cluster* optimal (*Elbow Method* dan *Silhouette Method*), serta metrik evaluasi *cluster* seperti *Silhouette Score*, *Davies-Bouldin Index* (DBI), dan *Calinski-Harabasz Index* (CH).

3. Bab 3 METODOLOGI PENELITIAN

Bab ini menjelaskan metodologi penelitian yang digunakan, meliputi tahap praproses data, pemodelan menggunakan algoritma *K-Means Clustering*, serta evaluasi *cluster*.

4. Bab 4 HASIL DAN DISKUSI

Bab ini menyajikan hasil penelitian yang meliputi implementasi algoritma *K-Means Clustering*, penentuan jumlah *cluster* optimal, serta hasil integrasi dengan karakteristik ekonomi wilayah berbasis data survei.

5. Bab 5 SIMPULAN DAN SARAN

Bab ini berisi simpulan yang menjawab rumusan masalah penelitian serta saran untuk penelitian selanjutnya.

