

BAB I

PENDAHULUAN

1.1 Latar Belakang

Ritel adalah toko penjualan barang Kebutuhan sehari-hari. Banyaknya pesaing di pasar, terutama di industri kosmetik memaksa para produsen dan pengembang supaya mencari cara untuk mengembangkan penjualan produk (Meta, 2020). Salah satu cara yaitu dengan memanfaatkan data transaksi. Data transaksi ini dapat diubah menjadi informasi yang menguntungkan untuk setiap toko di industri kosmetik, sistem komputerisasi digunakan oleh hampir setiap toko dan tempat penjualan untuk menyimpan data penjualan (Maryani et al., 2022). Data transaksi penjualan berisi banyak catatan transaksi penjualan dengan jumlah data yang sangat besar, itulah sebabnya jumlah data terus bertambah setiap hari (Kurniawati et al., 2021).

Jumlah data yang begitu besar akan menjadi masalah bagi toko jika tidak dimanfaatkan datanya. Meskipun sejumlah besar datanya dapat dikelola, tidak sedikit kemungkinan petunjuk untuk menggali data menjadi sebuah informasi. Biasanya data transaksi yang ada hanya digunakan sebagai arsip untuk ritel yang menimbulkan kendala dalam pemrosesan informasi, apabila data transaksi yang ada dalam database penjualan diabaikan, hal ini sebanding dengan penjual yang mengumpulkan data tanpa mengetahui apa manfaatnya di masa yang akan mendatang (Febrianti, 2021).

Permasalahan diatas terjadi pada toko KTT Kosmetik, sebagai suatu industri pengadaan barang kosmetik untuk konsumen. Dalam situasi ini pemrosesan data yang membantu bisnis untuk memperoleh informasi sangat dibutuhkan. Salah satu cara untuk mendapatkan informasi tersebut yaitu dengan menggunakan *data mining* untuk menganalisis data terkait transaksi penjualan yang terjadi di toko tersebut dan menarik kesimpulan terkait produk kosmetik yang paling diminati konsumen.

Metode *data mining* yang digunakan pada penelitian ini, untuk mengelompokkan jumlah penjualan barang yaitu metode *clustering*. Penerapan metode *clustering* dalam analisis penjualan menawarkan berbagai keuntungan yang signifikan serta memungkinkan pengelompokan data penjualan menjadi segmen-segmen kecil yang lebih homogen (PILLI SRI DURGA et al., 2023). Dengan penggunaan *clustering* dalam mengidentifikasi pola penjualan yang beragam dapat menentukan tren penjualan dan produk yang populer sehingga dapat menyusun strategi pemasaran yang lebih efektif dan efisien (Chenguang, 2022)

Terdapat empat algoritma yang umum digunakan pada *clustering* pengelompokan penjualan yaitu *K-Means Clustering*, *K-Medoids Clustering*, *DBSCAN (Density-Based Spatial Clustering of Applications with Noise)* dan *Hierarchical Clustering* (Mahmoud, 2021). Namun yang populer digunakan yaitu algoritma *K-means* dan algoritma *K-Medoids* (Prasan N.H, 2024). Pada penelitian ini, algoritma yang dilakukan adalah *K-Medoids*. Hal ini dilakukan karena algoritma *K-Medoids* memiliki keunggulan dalam menangani *outlier* dibandingkan dengan *K-Means* (Dyang et al., 2017) *K-Medoids* menggunakan objek yang paling representatif dari data sebagai pusat *cluster*, bukan rata-rata dari objek, sehingga lebih tahan terhadap nilai ekstrim atau *outlier* dengan representasi yang jelas membuat interpretasi dan analisis hasil *clustering* menjadi lebih mudah dan intuitif serta stabilitas cenderung lebih stabil dalam hasil *clustering* dibandingkan dengan algoritma seperti *K-Means* yang sensitif terhadap perubahan kecil pada data (Sindi et al., 2020). Penelitian terkait dengan penerapan algoritma *K-Means* maupun *K-Medoids* terdahulu yang telah banyak dilakukan yaitu dalam hal *clusterisasi* Data Bongkar Muat Provinsi Riau (Kamila et al., 2019) diperoleh kesimpulan bahwa dalam pengolahannya, algoritma *K-means* lebih cepat dan menghasilkan nilai pada *K-Means* yang dinilai lebih rendah baik dengan jumlah *cluster* $K=3$ dan $K=10$. Begitu juga pada penelitian yang telah dilakukan Untuk Pemetaan Hasil Produksi Buah-Buahan (Prasetyaningrum & Susanti, 2023) diperoleh hasil bahwa algoritma *K-Means* mendapatkan hasil nilai yang lebih kecil

dengan nilai 0,296 sedangkan hasil algoritma *K-Medoids* sebesar 0,507. Selain itu penelitian yang membandingkan algoritma *K-Means* dengan algoritma *K-Medoids Davies Bouldin Index* sebagai metode dalam analisis *cluster* menghasilkan nilai validitas sebesar 0,67 untuk *K-Means clustering* dan 1,78 untuk *K-Medoids* (Supriyadi et al., 2021).

Terdapat beberapa penelitian mengenai perbandingan penentuan jumlah κ optimal. Seperti yang dilakukan (Ramadhania, 2022) penetapan besaran *cluster* yang ideal pada setiap metode dipengaruhi oleh setiap entitas dan jumlah data skor rata-rata tertinggi yang diperoleh dengan menggunakan metode *Gap Statistic* adalah 0,125734. Kemudian penelitian (Ayu et al., 2019) adapun analisis perbandingan metode *Elbow* dan *Silhouette* pada algoritma *clustering K-Medoids* dalam pengelompokan produksi kerajinan Bali menggunakan evaluasi *cluster Davies Bouldin Index* (DBI).

Metode yang biasanya dipakai dalam penentuan jumlah *cluster* yang optimal, termasuk yang paling umum yaitu, menggunakan metode *Elbow*, metode *Silhouette Analysis*, dan *Gap Statistic*. Masing-masing metode memiliki kelebihan dan kekurangan, oleh karena itu perlu ketelitian dalam menggabungkan metode *clustering* yang digunakan, metode untuk menetapkan jumlah *cluster* yang optimal dan struktur data serta ukuran data (Ayu et al., 2019).

Berdasarkan penelitian sebelumnya, permasalahan yang dialami yaitu belum diterapkannya *data mining* dalam menemukan jenis barang yang paling banyak terjual. Oleh karena itu penulis tertarik untuk melakukan penelitian memakai teknik *data mining* menggunakan *clustering* algoritma *K-Medoids*, serta metode *Elbow*, *Silhouette Analysis*, dan *Gap Statistic* dengan judul penelitian **“Optimasi Jumlah Cluster Untuk Analisis Penjualan Barang Kosmetik Menggunakan *K-Medoids*”**.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan, maka rumusan masalah dalam penelitian ini yaitu:

1. Bagaimana menentukan jumlah *cluster* yang optimal untuk mengelompokkan data produk dalam menganalisa data transaksi penjualan kosmetik pada toko KTT?
2. Bagaimana memanfaatkan metode *K-Medoids* guna mempelajari dan menganalisa performa penjualan suatu produk?

1.3 Tujuan Penelitian

Berdasarkan latar belakang masalah dan rumusan masalah maka tujuan dari penelitian ini yaitu:

1. Melakukan analisis pencarian jumlah *cluster* optimal dengan mempertimbangkan perhitungan metode *Elbow*, *Silhouette Analysis*, dan *Gap Statistic*.
2. Mengelompokkan seluruh jenis produk berdasarkan kategori untuk selanjutnya melakukan analisis *clustering* menggunakan metode *K-Medoids* untuk mendapatkan hasil pengelompokkan jenis produk berdasarkan performa penjualannya.

1.4 Batasan Masalah

Terdapat beberapa Batasan masalah yang digunakan agar penelitian dapat dilakukan secara spesifik, adapun batasan masalah penelitian yang dilakukan adalah sebagai berikut:

1. Studi kasus yang digunakan untuk mengelompokkan data penjualan barang menggunakan data transaksi selama bulan Mei tahun 2024 pada toko KTT Kosmetik.
2. Data yang diambil hanya data pengelompokkan jumlah penjualan barang di toko KTT Kosmetik selama bulan Mei tahun 2024.
3. Data yang diolah dalam penelitian ini menggunakan metode *clustering* dengan menggunakan algoritma *K-Medoids*.
4. Penelitian ini sekedar menggunakan algoritma *K-Medoids* untuk menganalisis data penjualan tanpa membandingkan dengan metode *clustering* lain.
5. Jumlah *cluster* yang optimal akan ditentukan berdasarkan hasil dari metode *elbow*, *silhouette analysis*, dan *gap statistic* tanpa analisis komparatif terhadap metode lain.

6. *Tools* yang digunakan di dalam penerapan *data mining* ini adalah *Python*.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat bermanfaat bagi seluruh pihak yang terkait diantaranya:

1. Penelitian ini memiliki potensi untuk memberikan kontribusi yang signifikan bagi KTT Kosmetik dalam meningkatkan strategi pemasaran, manajemen stok, layanan pelanggan, dan efisiensi operasional.
2. Dengan menggunakan algoritma *K-Medoids*, dapat mengidentifikasi kelompok pelanggan dengan perilaku belanja yang serupa, sehingga memungkinkan strategi pemasaran yang lebih tepat sasaran.

1.6 Metodologi Penelitian

Metodologi penelitian merupakan Langkah-langkah yang dilakukan dalam penelitian ilmiah yang bertujuan untuk mendapatkan hasil sebagai tujuan dari penelitian tersebut dapat terpenuhi. Berikut ini adalah tahapan-tahapan penelitian diantaranya adalah sebagai berikut:

Teknik pengumpulan data

a. Observasi

Pendekatan dilakukan dengan mengamati secara langsung kejadian pada sistem jumlah penjualan jenis barang terlaris selama 1 bulan dalam 4 kali kunjungan, di KTT Kosmetik di Tasikmalaya.

b. Wawancara

Kegiatan ini dilakukan untuk mencari informasi mengenai jumlah penjualan barang terlaris KTT Kosmetik. Data yang dilakukan perhitungan *data mining* adalah jumlah penjualan barang terlaris di KTT Kosmetik di bulan Mei 2024.

c. Studi Kepustakaan

Pengumpulan data yang dilakukan secara langsung dari sumber-sumber lain seperti buku, skripsi, jurnal, dan dokumen lain yang berhubungan dengan penelitian.