

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Teknologi informasi dan komunikasi telah membawa perubahan signifikan dalam berbagai bidang kehidupan manusia, termasuk dalam sektor perdagangan. Dampak nyata dari perkembangan tersebut adalah munculnya perdagangan elektronik atau *e-commerce* yang semakin populer di kalangan masyarakat. *E-commerce* menawarkan kemudahan dalam bertransaksi karena konsumen tidak lagi harus datang langsung ke toko fisik. Selain itu, efisiensi waktu dan kemudahan akses menjadi faktor utama yang mendorong meningkatnya penggunaan *platform* belanja daring. Melalui toko *online*, konsumen dapat memilih dan membeli produk kapan saja dan dimana saja. Hal tersebut membuat proses transaksi menjadi lebih fleksibel dan praktis. Perubahan pola belanja ini juga menunjukkan adanya pergeseran perilaku konsumen ke arah digital. Oleh karena itu, *e-commerce* menjadi salah satu wujud transformasi digital yang memiliki peran penting dalam aktivitas ekonomi modern (Diskominfo Kalsel, 2020).

Meningkatnya jumlah transaksi pada *platform e-commerce* secara langsung menghasilkan data dalam volume yang sangat besar atau yang dikenal sebagai *big data*. Data tersebut merekam seluruh aktivitas pelanggan selama melakukan transaksi pembelian. Setiap transaksi menyimpan berbagai informasi penting, seperti waktu pembelian, jumlah produk yang dibeli, total nilai transaksi, serta jenis produk yang dipilih. Informasi ini memiliki nilai strategis apabila dianalisis secara tepat dan sistematis. Dengan pengolahan data yang baik, pelaku usaha dapat lebih paham mendalam mengenai kebiasaan belanja pelanggan. Selain itu, preferensi dan pola perilaku konsumen juga dapat diidentifikasi melalui data transaksi tersebut. Analisis data pelanggan menjadi dasar penting dalam perumusan strategi pemasaran yang lebih terarah dan personal. Namun, pada kenyataannya masih banyak toko *online* yang belum menerapkan segmentasi pelanggan secara jelas, sehingga strategi

pemasaran yang dijalankan cenderung bersifat umum dan kurang optimal (Olipian Resky, 2024).

Untuk memahami karakteristik dan perilaku pelanggan, maka metode yang sering digunakan adalah model *Recency, Frequency, Monetary (RFM)*. Model ini mengelompokkan pelanggan berdasar pada tiga indikator utama yang saling berkaitan. Indikator *recency* menggambarkan seberapa baru pelanggan terakhir kali melakukan transaksi. *Frequency* menunjukkan seberapa sering pelanggan melakukan pembelian dalam periode tertentu. Sementara itu, *monetary* mencerminkan total nilai transaksi yang dihasilkan oleh pelanggan. Dengan menggunakan ketiga indikator tersebut, pelaku usaha dapat mengklasifikasikan pelanggan secara lebih sistematis. Penerapan model RFM membantu dalam mengidentifikasi pelanggan potensial dan pelanggan yang kurang aktif. Segmentasi yang dihasilkan dapat digunakan untuk strategi pemasaran yang sesuai dengan karakteristik kelompok pelanggan. Oleh karena itu, model RFM menjadi pendekatan yang relevan dan efektif dalam mendukung pemasaran digital yang berbasis data (Tabianan & Velu, 2022).

Dalam proses segmentasi data pelanggan, salah satu teknik yang umum digunakan adalah *clustering*. *Clustering* merupakan metode pengelompokan data berdasarkan tingkat kemiripan karakteristik tertentu. Algoritma K-Means menjadi salah satu teknik *clustering* yang paling sering diterapkan karena kemampuannya dalam menangani data berukuran besar secara efisien. Keberhasilan algoritma K-Means sangat dipengaruhi pada metode perhitungan jarak yang digunakan. Setiap metode jarak memiliki karakteristik perhitungan yang berbeda, sehingga dapat menghasilkan pola *cluster* yang beragam. Pemilihan metode jarak yang kurang tepat dapat berdampak pada rendahnya kualitas hasil *clustering*. Penelitian sebelumnya menunjukkan bahwa penggunaan metode jarak yang berbeda menghasilkan nilai validitas *cluster* yang tidak sama. Studi oleh Buaton dan Solikhun (2023) menunjukkan bahwa metode Euclidean memberikan performa terbaik pada *dataset*

tertentu. Hal ini menegaskan bahwa tidak ada satu metode jarak yang selalu optimal untuk semua jenis data, sehingga pemilihan metode jarak perlu disesuaikan dengan karakteristik *dataset* yang dianalisis.

Penelitian ini bertujuan untuk membandingkan dua metode perhitungan jarak, yaitu *Chebyshev* dan *Canberra*, dalam algoritma K-Means. Fokus utama penelitian adalah mengelompokkan kebiasaan pelanggan berdasarkan data transaksi pada toko *online*. *Dataset* yang digunakan diperoleh dari *platform Kaggle* dan *UCI Machine Learning Repository*. Data tersebut terdiri dari sekitar 541.908 baris transaksi yang berasal dari sebuah toko *online* anonim. Atribut yang tersedia dalam *dataset* meliputi *InvoiceNo*, *StockCode*, *Description*, *Quantity*, *InvoiceDate*, *UnitPrice*, dan *Country*. Data transaksi ini kemudian diolah dan ditransformasikan ke dalam bentuk nilai RFM. Transformasi ini bertujuan untuk mempermudah analisis perilaku pelanggan.

Selain itu, karena jumlah data yang sangat besar dapat meningkatkan beban komputasi, memperlambat proses *preprocessing*, serta memperpanjang waktu eksekusi algoritma K-Means, maka dilakukan proses reduksi data menggunakan teknik *sampling*. Proses ini menghasilkan 4.339 data transaksi yang tetap merepresentasikan distribusi dan pola perilaku pelanggan secara keseluruhan. Penggunaan *sampling* bertujuan untuk menjaga keseimbangan antara efisiensi waktu pemrosesan dan kualitas hasil *clustering*, sehingga analisis tetap akurat namun lebih optimal secara komputasi. Melalui penelitian ini, diharapkan dapat diketahui metode jarak yang menghasilkan kualitas *cluster* terbaik berdasarkan nilai *Davies-Bouldin Index* serta efisiensi waktu eksekusi. Secara praktis, hasil segmentasi pelanggan ini dapat dimanfaatkan sebagai dasar pengambilan keputusan bisnis, seperti penentuan target promosi, pemberian program loyalitas kepada pelanggan bernilai tinggi, penyusunan strategi retensi bagi pelanggan kurang aktif, serta pengelolaan stok produk berdasarkan pola pembelian masing-masing segmen.

Dengan demikian, proses pengambilan keputusan tidak lagi berdasarkan perkiraan, melainkan berbasis data (*data-driven decision making*).

1.2 Rumusan Masalah

Didasari pada latar belakang yang sudah disebutkan sebelumnya, penelitian ini akan dirumuskan masalah sebagai berikut:

1. Bagaimana mengelompokkan kebiasaan pelanggan pada toko *online* menggunakan algoritma K-Means berbasis model RFM serta menentukan metode perhitungan jarak (*Chebyshev* atau *Canberra*) yang paling optimal berdasarkan waktu eksekusi dan nilai *Davies-Bouldin Index*?

1.3 Batasan Masalah

Dari rumusan masalah diatas maka dibuatlah batasan masalah agar penelitian yang dilakukan tidak melebar dari permasalahan yang dihadapi. Batasan masalah dalam penelitian ini yaitu:

1. Penelitian ini hanya berfokus pada analisis kebiasaan pelanggan berdasarkan transaksi pembelian, tanpa mempertimbangkan faktor eksternal seperti usia, jenis kelamin, atau lokasi pelanggan.
2. Data transaksi yang digunakan berasal dari *platform Kaggle* dan *UCI Machine Learning Repository*, yang dikumpulkan dari sebuah toko *online* anonim. Identitas toko tidak diungkapkan karena alasan kerahasiaan.
3. Transaksi pelanggan yang dianalisis telah ditransformasi menggunakan pendekatan *Recency*, *Frequency*, dan *Monetary (RFM)* sebagai dasar segmentasi pelanggan.
4. Algoritma *clustering* yang dipakai pada penelitian ini terbatas dengan algoritma k-means, dan tidak dilakukan perbandingan dengan algoritma *clustering* lainnya seperti DBSCAN atau Hierarchical *Clustering*.
5. Evaluasi hasil *clustering* hanya menggunakan dua metrik, yaitu *Davies-Bouldin Index* dan waktu eksekusi, tanpa menyertakan metrik lain seperti *Silhouette Score* atau *Dunn Index*.

6. Pengelompokan (*clustering*) akan dilakukan menggunakan software seperti *RapidMiner*.

1.4 Tujuan Penelitian

Berdasarkan batasan masalah yang ada, terdapat beberapa tujuan yang ingin dicapai. Adapun tujuan dari penelitian ini adalah:

1. Menerapkan Algoritma K-Means untuk mengelompokkan Kebiasaan Pelanggan pada Toko *Online* sesuai dengan Model RFM
2. Membandingkan metode perhitungan jarak *Chebyshev* dan *Canberra* pada algoritma k-mean untuk mengelompokkan Kebiasaan Pelanggan pada Toko *Online*.
3. Memberikan rekomendasi segmentasi pelanggan sebagai dasar pendukung pengambilan keputusan bisnis, seperti strategi promosi, program loyalitas, retensi pelanggan, dan perencanaan persediaan produk.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat yang luas baik secara akademis maupun praktis:

1. Manfaat Teoritis

Penelitian ini memberikan informasi penting dalam bidang *data mining*, terkait penerapan algoritma K-Means untuk menganalisis perilaku pelanggan pada toko *online*. Studi ini memperkaya pemahaman mengenai pemanfaatan teknik *clustering* dalam mengelompokkan pelanggan berdasarkan karakteristik transaksi. Selain itu, penelitian ini menyajikan gambaran komprehensif mengenai perbandingan metode perhitungan jarak *Chebyshev* dan *Canberra*. Perbandingan tersebut menunjukkan bagaimana masing-masing metode jarak memengaruhi efisiensi waktu eksekusi algoritma. Tidak hanya dari sisi waktu, penelitian ini juga menyoroti pengaruh metode jarak terhadap kualitas *cluster* yang

dihasilkan. Hasil analisis memberikan informasi yang relevan dalam menentukan metode jarak yang sesuai dengan karakteristik data pelanggan. Temuan dalam penelitian ini diharapkan dapat dijadikan rujukan bagi penelitian lebih lanjut. Dengan demikian, penelitian ini dapat mendukung pengembangan metode perhitungan jarak yang lebih optimal untuk analisis data pelanggan pada *platform e-commerce*.

2. Manfaat Praktis

Secara praktis, penelitian ini memberikan manfaat bagi pelaku usaha dalam mengelompokkan pelanggan berdasarkan pola perilaku belanja yang dimiliki. Pengelompokan tersebut memungkinkan penyusunan strategi pemasaran yang efektif dan tepat sasaran. Melalui segmentasi pelanggan yang lebih terstruktur, pelaku bisnis dapat meningkatkan kualitas perencanaan program loyalitas. Selain itu, personalisasi promosi dapat dilakukan secara lebih optimal sesuai dengan karakteristik masing-masing segmen pelanggan. Pengelolaan stok barang juga dapat disesuaikan dengan kebutuhan dan permintaan pelanggan pada setiap kelompok. Penelitian ini turut memberikan gambaran mengenai pemilihan metode *clustering* yang lebih efisien. Informasi tersebut membantu pelaku bisnis dalam menentukan pendekatan analisis data pelanggan yang paling sesuai. Dengan demikian, hasil penelitian ini dapat mendukung pengambilan keputusan strategis yang lebih akurat dan berbasis data.

1.6 Metode Penelitian

Penelitian ini menggunakan kuantitatif untuk menganalisis data kebiasaan pelanggan pada toko *online*. Hal yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Studi Literatur

Pengumpulan referensi yang relevan terkait algoritma K-Means, metode perhitungan jarak *Chebyshev* dan *Canberra*, serta konsep segmentasi

pelanggan menggunakan model *Recency, Frequency, Monetary* (RFM). Literatur diperoleh dari jurnal, buku, dan sumber daring terpercaya untuk mendukung pemahaman teoretis dan metodologis.

2. Pengolahan dan Transformasi Data

Data transaksi pelanggan toko *online* diolah dan dibersihkan terlebih dahulu. Kemudian dilakukan transformasi data mentah menjadi nilai RFM. Nilai *Recency, Frequency*, dan *Monetary* yang masing-masing diberikan skor skala 1-3 untuk proses *clustering*.

3. Implementasi *Clustering*

Proses *clustering* ini menggunakan algoritma *K-Means* dengan dua variasi metode jarak, yaitu *Chebyshev* dan *Canberra*. Implementasi dilakukan menggunakan perangkat lunak *RapidMiner* untuk mengelompokkan pelanggan berdasarkan nilai RFM.

4. Evaluasi dan Analisis Hasil

Hasil *clustering* dari masing-masing metode jarak dievaluasi menggunakan dua indikator: waktu eksekusi dan *Davies-Bouldin Index (DBI)*. Nilai *DBI* digunakan untuk menilai kualitas *cluster* berdasarkan kohesi dan separasi data. Hasil evaluasi dibandingkan untuk menentukan metode yang paling optimal.

1.7 Sistematika Penulisan

Penyusunan penelitian pada proposal skripsi ini dibagi atas beberapa bab yang akan dibahas secara rinci dengan sistematika penulisan sebagai berikut:

BAB 1 PENDAHULUAN

Bab ini berfungsi sebagai pengantar terhadap topik penelitian yang akan dibahas. Pada bagian ini dijelaskan latar belakang yang mendasari dilakukannya penelitian. Selain itu, permasalahan utama yang menjadi fokus penelitian juga diuraikan secara jelas. Bab ini turut memaparkan batasan-batasan masalah agar pembahasan tetap terarah. Tujuan yang dicapai melalui penelitian ini juga dijelaskan secara sistematis. Selanjutnya, manfaat yang diharapkan dari hasil

penelitian dipaparkan baik secara teoritis maupun praktis. Sebagai penutup, bab ini menyajikan sistematika penulisan untuk memberikan gambaran alur pembahasan pada bab selanjutnya.

BAB II TINJAUAN PUSTAKA

Bab ini memuat berbagai referensi yang bersumber dari penelitian-penelitian terdahulu. Penyusunan bab ini bertujuan untuk menunjukkan pemahaman penulis terhadap studi sebelumnya yang mempunyai keterkaitan dengan topik penelitian. Selain itu, kajian terhadap penelitian terdahulu digunakan sebagai dasar dalam memperkuat landasan teoritis penelitian. Bab ini juga berperan dalam mengidentifikasi posisi penelitian yang akan dilakukan. Dengan mengaitkan hasil penelitian sebelumnya, diharapkan terlihat kesinambungan antara penelitian terdahulu dan penelitian yang sedang dikembangkan. Hal ini membantu memperjelas kontribusi penelitian dalam bidang yang dikaji.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan secara detail metodologi penelitian yang diterapkan dalam pelaksanaan penelitian. Pembahasan mencakup tahapan pengumpulan data yang digunakan sebagai dasar analisis. Selain itu, teknik analisis data yang diterapkan juga dijelaskan secara sistematis. Prosedur penelitian yang dijalankan dipaparkan untuk menunjukkan alur pencapaian tujuan penelitian. Penjelasan metode dilakukan secara menyeluruh agar proses penelitian mudah dipahami. Dengan adanya pemaparan tersebut, penelitian diharapkan dapat direplikasi oleh peneliti lain. Hal ini juga memungkinkan hasil penelitian untuk diuji dan diverifikasi kembali secara ilmiah.

BAB IV HASIL DAN PEMBAHASAN

Bab ini memusatkan pembahasan pada hasil yang diperoleh dari pelaksanaan penelitian. Pada bagian ini disajikan hasil analisis data yang sudah dilakukan secara sistematis. Temuan-temuan utama penelitian dipaparkan secara jelas dan terstruktur. Selanjutnya, hasil penelitian tersebut dikaitkan dengan teori-teori yang relevan. Pembahasan dilakukan untuk melihat kesesuaian antara temuan penelitian dan landasan teoritis yang ada. Selain itu, bab ini juga memberikan penjelasan

mendalam mengenai makna dari hasil yang diperoleh. Dengan demikian, pembahasan dapat memberikan pemahaman yang komprehensif terhadap hasil penelitian ini.

BAB V PENUTUP

Bab ini menyajikan kesimpulan yang dirumuskan berdasarkan hasil pembahasan pada bab sebelumnya. Kesimpulan yang disusun menitikberatkan pada tingkat pencapaian tujuan penelitian yang telah ditetapkan. Selain itu, bab ini juga memuat saran sebagai bahan pertimbangan untuk pengembangan penelitian lebih lanjut. Rekomendasi yang diberikan diharapkan dapat menjadi acuan bagi penelitian di masa mendatang. Tidak hanya itu, saran yang disampaikan juga ditujukan bagi pihak yang berkepentingan terhadap hasil penelitian. Dengan demikian, bab ini berfungsi sebagai penutup yang merangkum hasil penelitian sekaligus memberikan arah pengembangan lebih lanjut.

DAFTAR PUSTAKA

Bagian ini memuat sumber rujukan yang digunakan dalam penyusunan skripsi. Referensi yang dicantumkan mencakup berbagai jenis sumber, seperti buku, jurnal, artikel ilmiah, serta sumber daring. Setiap sumber disusun secara sistematis sesuai dengan kaidah penulisan yang berlaku. Pencantuman daftar pustaka bertujuan untuk menunjukkan dasar teoritis dan empiris penelitian. Selain itu, daftar ini juga berfungsi sebagai bentuk penghargaan terhadap karya ilmiah yang dijadikan acuan. Dengan penyusunan yang sesuai format, bagian ini mendukung kejelasan dan kredibilitas penelitian.

