

BAB II

LANDASAN TEORI

2.1. Tingkat Penjualan

Tingkat penjualan adalah ukuran atau indikator yang menunjukkan seberapa banyak produk atau jasa berhasil dijual dalam periode waktu tertentu, baik harian, mingguan, bulanan, maupun tahunan [3]. Tingkat penjualan menjadi salah satu parameter penting dalam mengukur kinerja bisnis karena langsung berkaitan dengan pendapatan dan profitabilitas. Tingkat penjualan yang tinggi mencerminkan tingginya permintaan pasar terhadap produk, keberhasilan strategi pemasaran, serta efektivitas distribusi dan pelayanan [4]. Sebaliknya, penurunan tingkat penjualan dapat menjadi indikator adanya masalah, baik dari sisi kualitas produk, perubahan tren pasar, kompetisi, atau kurang efektifnya promosi yang dijalankan. Oleh karena itu, pemantauan tingkat penjualan secara rutin menjadi aspek krusial dalam kegiatan operasional bisnis.

Dalam konteks penelitian ini, tingkat penjualan di Toko Riski menjadi fokus utama yang dianalisis untuk melihat pola-pola pembelian pelanggan [5]. Dengan memanfaatkan metode *K-Means Clustering*, data penjualan dapat dikelompokkan ke dalam beberapa klaster, misalnya klaster pelanggan dengan tingkat pembelian tinggi, sedang, dan rendah. Pengelompokan ini membantu pemilik toko memahami segmen pelanggan yang paling berkontribusi terhadap peningkatan tingkat penjualan, serta mengidentifikasi kelompok pelanggan yang potensial untuk ditingkatkan loyalitasnya [6]. Dengan informasi ini, Toko Riski dapat

merancang strategi bisnis yang lebih terarah, seperti memberikan penawaran khusus kepada pelanggan setia atau menyusun promosi menarik untuk menarik pelanggan dengan tingkat pembelian rendah agar menjadi lebih aktif berbelanja.

2.2. Penjualan Pakaian

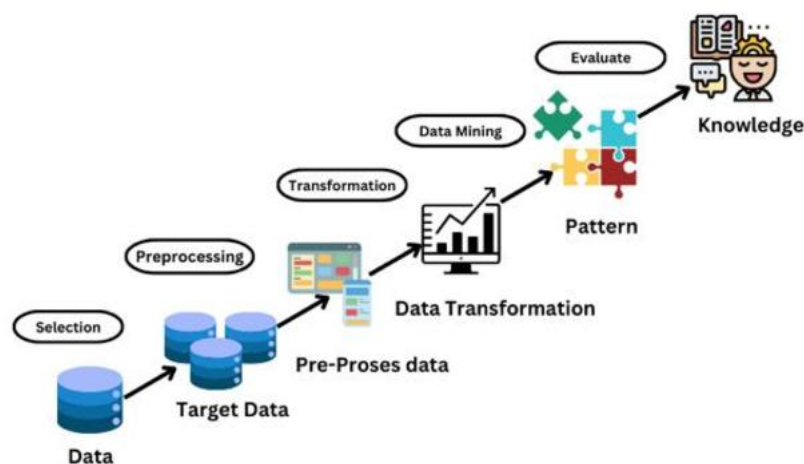
Penjualan pakaian merupakan salah satu sektor bisnis yang memiliki tingkat persaingan tinggi dan dinamis, terutama karena tren mode yang terus berubah seiring waktu. Tingkat penjualan pakaian sangat dipengaruhi oleh berbagai faktor, seperti musim, tren fashion, harga, kualitas produk, strategi promosi, hingga loyalitas pelanggan. Di banyak daerah, usaha penjualan pakaian — mulai dari skala besar seperti department store hingga toko kecil di lingkungan lokal — selalu mengalami fluktuasi penjualan yang signifikan, terutama saat menghadapi momen-momen khusus seperti Ramadan, Lebaran, Natal, atau pergantian tahun. Oleh karena itu, pemilik usaha harus mampu membaca pola penjualan dengan cermat agar dapat menjaga keseimbangan stok barang dan memaksimalkan keuntungan.

Dalam penelitian ini, fokus diarahkan pada analisis data penjualan pakaian di Toko Riski sebagai upaya memahami pola pembelian pelanggan yang berpengaruh pada tingkat penjualan. Dengan menggunakan metode *K-Means Clustering*, penelitian ini bertujuan untuk mengelompokkan pelanggan ke dalam beberapa klaster berdasarkan perilaku belanja mereka, sehingga pemilik toko dapat merancang strategi yang lebih spesifik untuk setiap kelompok. Misalnya, pelanggan yang sering membeli produk diskon dapat diarahkan pada promosi khusus, sementara pelanggan loyal dengan pembelian rutin dapat diberikan

program reward. Dengan demikian, pemanfaatan data penjualan melalui pendekatan *Data Mining* ini diharapkan mampu meningkatkan efektivitas strategi pemasaran dan mendukung peningkatan tingkat penjualan secara keseluruhan.

2.3. *Knowledge Discovery in Database*

Knowledge Discovery in Databases (KDD) adalah suatu proses sistematis untuk menemukan pengetahuan yang berguna dari kumpulan data yang besar [7]. KDD mencakup serangkaian tahapan yang saling terkait, yaitu: pemilihan data (selection), pra-pemrosesan (preprocessing), transformasi data, *Data Mining*, dan evaluasi atau interpretasi hasil. Proses ini bertujuan untuk mengekstrak informasi tersembunyi yang sebelumnya tidak diketahui namun memiliki nilai strategis bagi pengambilan keputusan [8]. *Data Mining* merupakan inti dari proses KDD, yaitu tahap di mana algoritma atau teknik tertentu seperti klasifikasi, *Clustering*, atau asosiasi diterapkan untuk menemukan pola atau hubungan dalam data [9]. Namun, KDD tidak hanya berfokus pada penerapan algoritma, melainkan pada keseluruhan proses yang memastikan data siap untuk dianalisis hingga hasil yang diperoleh dapat diinterpretasikan dengan baik [10].



Gambar 2. 1. Knowledge Discovery in Database

1. Selection (Seleksi Data), Tahap ini bertujuan untuk memilih data yang relevan dari basis data yang tersedia. Tidak semua data yang dimiliki bersifat penting, sehingga perlu dipilih atribut-atribut yang sesuai dengan tujuan analisis. Dalam konteks penelitian penjualan pakaian, data seperti jumlah transaksi, total pembelian, dan metode pembayaran merupakan bagian yang dipilih untuk dianalisis lebih lanjut.
2. Preprocessing (Prapemrosesan Data), Pada tahap ini, data yang telah dipilih dibersihkan dari kesalahan, data kosong, atau data yang tidak konsisten. Prapemrosesan juga meliputi normalisasi data, konversi atribut kategorikal ke bentuk numerikal, serta penghapusan duplikasi. Tujuannya adalah memastikan kualitas data agar hasil analisis lebih akurat dan dapat diandalkan.
3. Transformation (Transformasi Data), Transformasi data dilakukan dengan mengubah data mentah ke dalam format yang sesuai untuk proses *Data Mining*. Ini bisa berupa pengkodean data, normalisasi nilai numerik, atau pembuatan variabel baru dari data yang sudah ada. Dalam penelitian ini, transformasi dilakukan misalnya dengan mengubah "tinggi/rendah" menjadi nilai numerik dan melakukan normalisasi agar setiap atribut memiliki bobot yang seimbang.
4. *Data Mining*, merupakan inti dari proses KDD, yaitu penerapan metode atau algoritma untuk menemukan pola atau informasi tersembunyi dalam data. Dalam penelitian ini, digunakan algoritma *K-Means Clustering* untuk mengelompokkan pelanggan berdasarkan karakteristik belanja mereka. Hasil

dari tahap ini adalah pembentukan klaster yang memiliki pola kemiripan internal.

5. Interpretation/Evaluation (Interpretasi dan Evaluasi), Tahap akhir ini bertujuan untuk mengevaluasi hasil dari proses *Data Mining* dan menginterpretasikannya dalam konteks bisnis atau penelitian. Pola atau klaster yang ditemukan dianalisis untuk melihat apakah hasil tersebut bermanfaat dan dapat digunakan untuk pengambilan keputusan. Misalnya, klaster yang terbentuk bisa dijadikan dasar untuk strategi pemasaran atau pengelompokan pelanggan potensial.

Dalam konteks penelitian ini, proses KDD digunakan sebagai kerangka kerja untuk menganalisis data penjualan pakaian di Toko Riski. Data transaksi yang telah terkumpul terlebih dahulu akan melalui tahapan seleksi dan pembersihan (preprocessing), seperti menghapus data duplikat atau data yang tidak relevan. Setelah itu, data akan ditransformasikan ke dalam bentuk yang sesuai untuk proses *Clustering*. Pada tahap *Data Mining*, algoritma *K-Means Clustering* akan diterapkan untuk mengelompokkan pelanggan berdasarkan pola pembelian mereka. Hasil dari pengelompokan ini kemudian dianalisis dan diinterpretasikan untuk membantu pemilik toko dalam memahami karakteristik setiap klaster pelanggan, sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan bisnis yang lebih efektif dan efisien.

2.3.1. Data Mining

Data Mining merupakan proses ekstraksi atau penggalian informasi yang bermanfaat dan tersembunyi dari sejumlah besar data [11]. Proses ini melibatkan teknik statistik, matematika, dan teknologi komputer untuk menemukan pola,

hubungan, atau tren tertentu dalam data yang tidak terlihat secara kasat mata [12]. *Data Mining* sering digunakan dalam berbagai bidang seperti bisnis, kesehatan, pendidikan, dan keuangan untuk mendukung pengambilan keputusan yang lebih cerdas. Beberapa metode populer dalam *Data Mining* meliputi klasifikasi, asosiasi, regresi, dan *Clustering* [13]. *Clustering* sendiri merupakan teknik untuk mengelompokkan data ke dalam kelompok-kelompok (klaster) berdasarkan kemiripan karakteristik, tanpa adanya label atau kategori yang sudah ditentukan sebelumnya (unsupervised learning) [14].

Dalam konteks penelitian ini, *Data Mining* diterapkan untuk menganalisis data transaksi penjualan pakaian di Toko Riski dengan menggunakan metode *K-Means Clustering*, salah satu teknik pengelompokan data yang paling umum digunakan [15]. Metode ini memungkinkan peneliti untuk membagi data pelanggan berdasarkan pola pembelian mereka ke dalam beberapa klaster, sehingga dapat diperoleh gambaran yang lebih jelas mengenai karakteristik masing-masing kelompok pelanggan [16]. Dengan pendekatan ini, Toko Riski dapat mengoptimalkan strategi bisnisnya melalui pemahaman yang lebih baik terhadap preferensi pelanggan, pengelolaan stok barang yang lebih efisien, serta penetapan strategi promosi yang tepat sasaran. Penerapan *Data Mining* diharapkan menjadi solusi praktis dan efisien bagi toko dalam memanfaatkan data yang selama ini belum dimaksimalkan secara analitis.

2.3.2. Database dan Data Processing

Dalam dunia bisnis modern, database memegang peranan penting dalam menyimpan, mengelola, dan merekam setiap transaksi penjualan yang terjadi [17]. Terutama pada usaha retail seperti penjualan pakaian, database membantu

mencatat detail seperti waktu transaksi, jenis produk yang terjual, jumlah pembelian, hingga informasi pelanggan [18]. Data ini, jika diolah dengan baik, dapat memberikan gambaran yang jelas mengenai tren penjualan, produk-produk yang paling diminati, periode penjualan tertinggi, serta pola-pola khusus lainnya yang tidak mudah terlihat hanya dari pengamatan manual [19]. Proses *data processing* atau pengolahan data kemudian menjadi tahap krusial untuk membersihkan, menyaring, dan menyiapkan data agar siap dianalisis, sehingga informasi yang dihasilkan lebih akurat dan relevan untuk mendukung keputusan bisnis [20].

Dalam penelitian ini, database penjualan di Toko Riski akan menjadi sumber utama yang digunakan untuk melakukan proses *Data Mining* dengan metode *K-Means Clustering* [21]. Data penjualan yang telah diproses akan digunakan untuk mengelompokkan pelanggan atau transaksi berdasarkan pola pembelian mereka. Tahap *data processing* akan membantu memastikan bahwa data yang digunakan bebas dari duplikasi, kesalahan pencatatan, atau data kosong yang bisa memengaruhi akurasi hasil *Clustering* [22]. Dengan pendekatan ini, penelitian dapat menghasilkan klaster yang benar-benar mencerminkan perilaku pelanggan, yang pada akhirnya membantu pemilik Toko Riski dalam merumuskan strategi penjualan yang lebih terarah, seperti pengelolaan stok barang, program loyalitas, atau promosi khusus untuk kelompok pelanggan tertentu.

2.3.3. Visualization

Visualisasi data adalah teknik penting dalam analisis tingkat penjualan karena memungkinkan informasi kompleks disajikan dalam bentuk grafis yang mudah dipahami [23]. Dengan menggunakan grafik batang, diagram garis, pie

chart, atau heatmap, pemilik usaha dapat melihat pola penjualan dari waktu ke waktu, mengidentifikasi produk dengan penjualan tertinggi atau terendah, serta mendeteksi tren musiman yang mungkin memengaruhi performa bisnis. Visualisasi ini membantu dalam mengenali fluktuasi penjualan yang tidak terlihat jika hanya mengandalkan data angka mentah, sehingga mendukung pengambilan keputusan yang lebih cepat dan tepat [24].

Dalam penelitian ini, visualisasi data digunakan untuk menampilkan hasil pengelompokan pelanggan berdasarkan metode *K-Means Clustering* pada data penjualan pakaian di Toko Riski [25]. Dengan visualisasi klaster yang dihasilkan, peneliti dapat mempermudah interpretasi pola belanja antar kelompok pelanggan, misalnya klaster pelanggan dengan frekuensi belanja tinggi, menengah, atau rendah. Hasil visualisasi ini tidak hanya membantu pemilik toko memahami karakteristik tiap kelompok, tetapi juga menjadi dasar untuk menyusun strategi pemasaran, pengaturan stok barang, dan penawaran promo yang disesuaikan dengan kebutuhan masing-masing segmen pelanggan.

2.3.4. Statistik

Statistik merupakan cabang ilmu yang berfokus pada pengumpulan, pengolahan, analisis, interpretasi, dan penyajian data untuk membantu dalam pengambilan keputusan yang lebih tepat [26]. Dalam konteks tingkat penjualan, statistik sangat penting digunakan untuk mengukur performa usaha dari waktu ke waktu, mengevaluasi pola pembelian pelanggan, serta memprediksi tren di masa mendatang [27]. Melalui teknik statistik deskriptif seperti rata-rata, median, dan persentase, pemilik usaha dapat mengetahui produk apa saja yang paling laku, kapan periode penjualan tertinggi terjadi, serta seberapa besar variasi dalam

permintaan barang. Sementara itu, statistik inferensial dapat membantu memperkirakan atau menguji dugaan mengenai perilaku pembeli berdasarkan sampel data.

Dalam penelitian ini, statistik digunakan sebagai dasar untuk memahami pola dan tren penjualan pakaian di Toko Riski sebelum diterapkan metode *Data Mining* seperti *K-Means Clustering* [28]. Analisis statistik membantu peneliti menggali informasi penting dari data transaksi, seperti frekuensi pembelian pelanggan, jenis produk favorit, dan variasi penjualan antar waktu [29]. Dengan demikian, hasil statistik menjadi fondasi awal yang mendukung proses *Clustering*, karena memberikan gambaran menyeluruh tentang data yang akan dikelompokkan. Integrasi antara statistik dan *Data Mining* memungkinkan peneliti menghasilkan wawasan yang lebih komprehensif dan mendalam, sehingga rekomendasi yang dihasilkan lebih tepat sasaran untuk meningkatkan strategi bisnis Toko Riski.

2.3.5. Pattern Recognition

Pattern Recognition atau pengenalan pola merupakan salah satu cabang penting dalam bidang ilmu komputer dan *Data Mining* yang bertujuan untuk mengidentifikasi dan mengekstrak pola-pola tertentu dari sekumpulan data [30]. Dalam konteks tingkat penjualan, *Pattern Recognition* berperan besar dalam menemukan tren, kecenderungan, atau perilaku yang berulang, seperti pola kenaikan atau penurunan permintaan, pola musiman, serta segmentasi pelanggan berdasarkan preferensi pembelian. Dengan mengenali pola-pola tersebut, pemilik usaha dapat memprediksi kebutuhan pasar, mengoptimalkan stok, serta menyusun

strategi pemasaran yang lebih efektif, sehingga performa bisnis dapat meningkat secara signifikan.

Dalam penelitian ini, konsep *Pattern Recognition* diterapkan melalui penggunaan metode *K-Means Clustering* untuk mengenali pola pembelian pelanggan di Toko Riski [31]. Dengan menganalisis data transaksi penjualan menggunakan teknik *Clustering*, peneliti berusaha menemukan kelompok-kelompok pelanggan yang memiliki pola pembelian serupa, seperti pelanggan yang sering membeli produk diskon, pelanggan loyal dengan pembelian rutin, atau pelanggan musiman. Hasil pengenalan pola ini tidak hanya membantu pemilik toko memahami perilaku konsumen secara lebih mendalam, tetapi juga menjadi dasar untuk merumuskan strategi bisnis yang berbasis data, seperti penawaran personalisasi, penataan stok yang lebih efisien, dan kampanye promosi yang tepat sasaran.

2.4. Metode *K-Means*

Metode *K-Means* merupakan salah satu algoritma unsupervised learning yang paling populer dan banyak digunakan dalam proses *Clustering* atau pengelompokan data [32]. *K-Means* bekerja dengan cara membagi sekumpulan data ke dalam sejumlah klaster (kelompok) berdasarkan kemiripan karakteristik antar data. Setiap klaster memiliki pusat gravitasi yang disebut *centroid*, dan proses pembagian data didasarkan pada jarak terdekat masing-masing data ke centroid. Algoritma ini sangat cocok digunakan ketika tujuan analisis adalah untuk menemukan pola atau segmentasi dalam kumpulan data yang tidak memiliki label atau target output.

$$d(x_i, c_j) = \sqrt{\sum_{k=1}^m (x_{ik} - c_{jk})^2}$$

Keterangan:

- x_i : titik data ke- i
- c_j : pusat cluster ke- j
- m : jumlah atribut atau dimensi data

Pada awalnya, pengguna harus menentukan jumlah kluster (K) yang diinginkan. Kemudian, algoritma *K-Means* secara acak memilih K titik sebagai centroid awal [33]. Setiap data akan dihitung jaraknya ke seluruh centroid, dan data tersebut akan dimasukkan ke dalam kluster yang memiliki jarak terdekat. Setelah semua data dikelompokkan, algoritma akan menghitung ulang posisi centroid berdasarkan rata-rata data dalam masing-masing kluster. Proses ini diulang terus hingga tidak ada lagi perubahan signifikan dalam pembagian kluster, atau hingga tercapai jumlah iterasi maksimum.

Kelebihan utama dari *K-Means* adalah kesederhanaannya, efisiensi waktu komputasi, dan kemampuannya untuk menangani dataset berukuran besar dengan atribut numerik. Namun, *K-Means* juga memiliki beberapa keterbatasan, seperti sensitif terhadap pemilihan centroid awal dan ketergantungannya pada jumlah kluster yang ditentukan di awal. Oleh karena itu, pemilihan jumlah K yang tepat menjadi aspek penting dalam keberhasilan proses *Clustering* menggunakan metode ini. Beberapa metode seperti Elbow Method atau Silhouette Score sering digunakan untuk membantu menentukan nilai K yang optimal.

Dalam konteks penelitian penjualan pakaian di Toko Riski, metode *K-Means* digunakan untuk mengelompokkan pelanggan berdasarkan atribut-atribut seperti jumlah transaksi, total pembelian, dan metode pembayaran. Dengan *Clustering* ini, peneliti dapat memahami perilaku pelanggan berdasarkan pola belanja mereka, sehingga dapat digunakan untuk perencanaan strategi pemasaran yang lebih efektif. Hasil dari *K-Means* dapat dijadikan dasar untuk memberikan pelayanan yang lebih personal, promosi yang tepat sasaran, serta pengelolaan stok barang yang lebih akurat sesuai kebutuhan masing-masing kelompok pelanggan.

2.4.1. Fungsi Metode *K-Means*

Pada metode *K-Means*, terdapat beberapa fungsi metode *K-Means* yang dapat di implementasikan yaitu sebagai berikut.

1. Segmentasi Data, *K-Means* berfungsi untuk membagi data besar menjadi kelompok-kelompok kecil berdasarkan kesamaan karakteristik. Dalam bisnis, segmentasi ini bisa digunakan untuk membagi pelanggan, produk, atau wilayah penjualan berdasarkan pola yang muncul dari data.
2. Menganalisis Pola, Dengan membentuk klaster berdasarkan atribut tertentu, *K-Means* membantu dalam mengungkap pola tersembunyi dalam dataset yang sebelumnya tidak terlihat. Ini sangat berguna untuk pengambilan keputusan berbasis data.
3. Reduksi Kompleksitas Data, *Clustering* dengan *K-Means* menyederhanakan kumpulan data besar menjadi beberapa kelompok, sehingga mempermudah analisis lanjutan seperti visualisasi atau prediksi pada masing-masing kelompok secara terpisah.

4. Dasar untuk Strategi Bisnis, Hasil dari *Clustering* dapat dijadikan pijakan dalam membuat strategi bisnis, seperti penentuan promosi, manajemen stok, hingga pengembangan produk berdasarkan karakteristik masing-masing klaster.

2.4.2. Model Metode *K-Means*

Pada penelitian ini, terdapat model yang dapat digunakan pada penelitian ini. Model ini merupakan model yang ada pada metode *K-Means* yaitu sebagai berikut.

1. Centroid-Based *Clustering*, Model ini berfokus pada representasi setiap klaster dengan satu titik pusat (centroid). Setiap data ditempatkan ke dalam klaster yang memiliki centroid terdekat berdasarkan jarak Euclidean atau metrik jarak lainnya.
2. Iterative Refinement Model, *K-Means* termasuk dalam model iteratif karena prosesnya melibatkan pengulangan: mengelompokkan data, menghitung centroid baru, dan mengulangi proses hingga tidak ada perubahan signifikan. Ini menjadikan *K-Means* efisien dalam menemukan struktur alami dalam data.
3. Partitioning Method, Dalam model ini, *K-Means* bekerja dengan membagi dataset ke dalam K bagian terpisah yang tidak tumpang tindih. Setiap data hanya bisa berada dalam satu klaster, yang membuatnya sederhana namun efektif untuk banyak aplikasi praktis.

2.5. Model *Clustering*

Clustering adalah salah satu teknik dalam *Data Mining* yang bertujuan untuk mengelompokkan sekumpulan data ke dalam beberapa grup atau klaster berdasarkan kemiripan karakteristik antar data [34]. Metode ini bersifat unsupervised learning, artinya tidak memerlukan label atau kategori sebelumnya dalam proses pengelompokan. Salah satu algoritma *Clustering* yang paling populer dan banyak digunakan adalah *K-Means Clustering* [35]. Algoritma ini bekerja dengan cara membagi data menjadi sejumlah klaster yang telah ditentukan sebelumnya (k), lalu mengelompokkan data berdasarkan jarak terdekat ke pusat klaster (centroid). Proses ini berulang hingga posisi centroid stabil dan klaster tidak berubah lagi, sehingga menghasilkan pembagian data yang optimal sesuai pola yang ditemukan.

Dalam konteks penelitian ini, model *K-Means Clustering* digunakan untuk mengelompokkan data penjualan pakaian di Toko Riski berdasarkan pola pembelian pelanggan. Dengan menerapkan metode ini, diharapkan dapat ditemukan kelompok pelanggan dengan karakteristik dan preferensi yang berbeda, sehingga toko dapat memahami perilaku konsumen secara lebih terperinci. Hasil *Clustering* tersebut nantinya akan menjadi dasar dalam merancang strategi pemasaran, pengelolaan stok, dan pelayanan yang lebih tepat sasaran, sehingga meningkatkan efisiensi operasional dan daya saing Toko Riski di pasar yang kompetitif. Pendekatan ini memungkinkan pemanfaatan data transaksi yang selama ini kurang dimaksimalkan menjadi informasi strategis yang bernilai tinggi.

2.6. Alat Bantu Program Aplikasi Orange

Orange adalah sebuah platform open-source yang dirancang untuk keperluan analisis data dan *Data Mining* dengan antarmuka yang user-friendly dan berbasis visualisasi. Aplikasi ini menyediakan berbagai fitur lengkap untuk proses eksplorasi data, pemodelan, dan evaluasi hasil, termasuk algoritma *Clustering* seperti *K-Means*, klasifikasi, regresi, serta visualisasi interaktif yang memudahkan pengguna memahami pola data secara intuitif. Orange sangat populer digunakan di kalangan akademisi dan praktisi karena kemampuannya dalam mengolah data tanpa memerlukan keahlian pemrograman yang mendalam. Selain itu, Orange juga memungkinkan integrasi dengan bahasa pemrograman Python untuk analisis yang lebih kompleks.

Dalam konteks penelitian ini, aplikasi Orange dipilih sebagai alat bantu utama untuk menerapkan metode *K-Means Clustering* pada data penjualan pakaian di Toko Riski. Dengan menggunakan Orange, proses pengelompokan data transaksi dapat dilakukan secara efisien dan visual, sehingga memudahkan peneliti dalam mengidentifikasi pola pembelian pelanggan dan membentuk kluster-kluster yang relevan. Keunggulan Orange dalam menyajikan hasil analisis dalam bentuk grafik dan diagram interaktif juga mendukung pemahaman yang lebih baik terhadap karakteristik masing-masing kelompok pelanggan, sehingga hasil penelitian ini dapat diinterpretasikan dengan lebih jelas dan diaplikasikan secara praktis oleh pemilik toko.

2.7. Kelebihan dan Kekurangan Metode *K-Means*

2.7.1. Kelebihan Metode *K-Means*

Salah satu kelebihan utama dari metode *K-Means* adalah kesederhanaannya dalam implementasi. Algoritma ini relatif mudah dipahami dan tidak memerlukan

perhitungan matematis yang rumit, sehingga sangat cocok digunakan oleh peneliti atau praktisi yang baru memulai eksplorasi dalam bidang *Data Mining*. Dengan hanya memerlukan beberapa langkah seperti inisialisasi centroid, perhitungan jarak, dan pembaruan centroid, algoritma ini dapat digunakan secara praktis dalam berbagai aplikasi *Clustering*, termasuk untuk analisis pelanggan, segmentasi pasar, dan pengelompokan dokumen.

Selain sederhana, metode *K-Means* juga memiliki keunggulan dari segi efisiensi komputasi. *K-Means* memiliki kompleksitas waktu yang rendah dibandingkan dengan algoritma *Clustering* lainnya seperti DBSCAN atau hierarchical *Clustering*, terutama saat digunakan pada dataset berukuran besar. Proses iteratif dalam *K-Means* biasanya konvergen dengan cepat dalam beberapa langkah, terutama ketika jumlah klaster (K) tidak terlalu besar. Oleh karena itu, algoritma ini sangat cocok diterapkan untuk kebutuhan bisnis yang memerlukan pemrosesan data secara cepat dan real-time.

Kelebihan lainnya terletak pada kemampuannya dalam menghasilkan klaster yang bersifat eksklusif dan saling terpisah. Dengan kata lain, setiap data hanya akan masuk ke satu klaster tertentu, sehingga interpretasi hasil menjadi lebih mudah. Hal ini sangat berguna ketika tujuannya adalah untuk menyusun strategi berdasarkan kelompok pelanggan yang berbeda-beda. Misalnya, dalam kasus toko pakaian, klaster pelanggan dapat digunakan untuk menentukan strategi diskon, stok barang, dan pendekatan promosi sesuai karakteristik masing-masing kelompok.

K-Means juga sangat fleksibel dalam penggunaannya dan dapat dengan mudah dikombinasikan dengan teknik lain, seperti analisis visualisasi atau algoritma supervised learning untuk keperluan lanjutan. Selain itu, *K-Means* bekerja sangat baik pada data berdimensi rendah dengan distribusi yang cukup homogen, menjadikannya solusi efektif untuk analisis eksploratif. Kemampuannya dalam menangkap pola umum membuat *K-Means* menjadi salah satu metode *Clustering* yang masih banyak digunakan hingga saat ini dalam berbagai sektor seperti e-commerce, perbankan, pendidikan, dan lainnya.

2.7.2. Kekurangan Metode *K-Means*

Namun demikian, metode *K-Means* juga memiliki beberapa kekurangan yang perlu diperhatikan. Salah satu kelemahan utamanya adalah ketergantungannya pada jumlah klaster (K) yang harus ditentukan di awal. Jika jumlah K yang dipilih tidak sesuai dengan struktur alami data, maka hasil *Clustering* yang diperoleh bisa saja tidak optimal atau bahkan menyesatkan. Dalam banyak kasus, peneliti harus mencoba beberapa nilai K dan mengevaluasi hasilnya dengan teknik tambahan seperti *Elbow Method* atau *Silhouette Score*, yang dapat memakan waktu dan sumber daya.

Selain itu, algoritma *K-Means* sangat sensitif terhadap pemilihan centroid awal. Jika centroid awal yang dipilih terlalu dekat atau jauh satu sama lain secara tidak merata, maka hasil akhir klaster yang terbentuk bisa tidak konsisten atau bahkan terjebak pada solusi lokal yang bukan merupakan hasil terbaik secara global. Untuk mengatasi hal ini, beberapa versi pengembangan *K-Means* seperti

K-Means++ telah dibuat, yang bertujuan untuk memilih titik awal secara lebih cerdas agar konvergensi menuju solusi yang lebih baik.

Kekurangan lainnya terletak pada ketidaksesuaian *K-Means* untuk menangani data berdimensi tinggi atau data dengan bentuk klaster yang kompleks, seperti klaster yang berbentuk elips, tidak simetris, atau saling tumpang tindih. Karena *K-Means* menggunakan jarak Euclidean sebagai dasar pengelompokan, algoritma ini bekerja optimal hanya untuk data yang memiliki distribusi bulat dan homogen. Ketika data memiliki variabilitas bentuk yang tinggi atau noise yang besar, *K-Means* dapat gagal dalam mengidentifikasi batas antar klaster secara akurat.

Terakhir, *K-Means* tidak cocok untuk data kategorikal secara langsung, karena algoritma ini hanya dapat menangani data numerik yang dapat dihitung jaraknya. Oleh karena itu, jika data berisi atribut kategorikal seperti "jenis kelamin", "metode pembayaran", atau "kategori produk", maka diperlukan proses transformasi terlebih dahulu seperti konversi label ke angka atau one-hot encoding. Proses ini tidak hanya menambah langkah tambahan dalam preprocessing, tetapi juga berisiko mengurangi interpretabilitas hasil akhir, terutama jika transformasi tidak dilakukan dengan hati-hati.

2.8. Evaluasi Model *K-Means*

Evaluasi terhadap model *K-Means* dapat dilakukan dengan mengamati sejauh mana model mampu mengelompokkan data secara akurat dan bermakna. Salah satu cara paling umum untuk mengevaluasi hasil *Clustering* adalah dengan melihat *intra-cluster similarity* dan *inter-cluster dissimilarity*, yaitu kemiripan

data dalam satu klaster dan perbedaan antara klaster satu dengan yang lain. Jika model *K-Means* bekerja dengan baik, maka data yang berada dalam satu kelompok akan memiliki karakteristik yang seragam, sedangkan antar kelompok akan tampak perbedaan yang signifikan. Dalam praktiknya, metrik seperti *Silhouette Score*, *Within Cluster Sum of Squares (WCSS)*, dan *Davies-Bouldin Index* sering digunakan untuk mengukur kualitas hasil *Clustering*. Nilai yang lebih baik menunjukkan bahwa model mampu membentuk klaster yang kompak dan saling terpisah dengan baik.

Selain metrik numerik, evaluasi model *K-Means* juga dapat dilakukan dengan cara visual, terutama jika data memiliki dua atau tiga dimensi. Visualisasi menggunakan scatter plot dengan warna berbeda untuk masing-masing klaster dapat membantu melihat apakah pola pengelompokan sudah sesuai harapan. Jika titik-titik data tersebar rapi membentuk beberapa kelompok yang tidak saling tumpang tindih, maka hasil *Clustering* dianggap memuaskan. Namun, jika terdapat banyak tumpang tindih antar klaster atau penyebaran titik yang tidak jelas, ini bisa menjadi indikasi bahwa pemilihan jumlah klaster (*K*) kurang tepat atau data mengandung noise tinggi. Evaluasi visual ini juga sangat berguna untuk interpretasi hasil, terutama bagi pemangku kepentingan non-teknis yang ingin memahami segmentasi data secara intuitif.

Lebih lanjut, evaluasi juga dapat dilihat dari sisi kegunaan atau manfaat hasil *Clustering* dalam konteks aplikasi nyata. Dalam kasus penelitian penjualan pakaian di Toko Riski, model *K-Means* dievaluasi berdasarkan apakah klaster yang dihasilkan mampu merepresentasikan segmentasi pelanggan secara logis—

misalnya, apakah pelanggan yang sering bertransaksi dan melakukan pembelian dalam jumlah besar terkumpul dalam satu klaster khusus. Apabila hasil klaster sesuai dengan tujuan bisnis seperti peningkatan layanan, strategi promosi personalisasi, atau efisiensi stok barang, maka model dapat dianggap efektif. Evaluasi semacam ini menekankan bahwa validitas model tidak hanya dilihat dari angka, tetapi juga dari sejauh mana hasil tersebut bisa diinterpretasikan dan dimanfaatkan secara praktis dalam pengambilan keputusan.

2.9. Kelebihan Penelitian

Penelitian ini memiliki kelebihan utama dalam penerapan metode *Data Mining*, khususnya algoritma *K-Means Clustering*, yang dapat menggali informasi tersembunyi dari data transaksi penjualan yang sebelumnya belum dimanfaatkan secara optimal. Melalui pendekatan ini, data pelanggan yang awalnya hanya berupa angka dan catatan transaksi, dapat diolah menjadi kelompok-kelompok dengan karakteristik tertentu. Hal ini memberikan nilai tambah bagi Toko Riski karena hasil pengelompokan dapat digunakan sebagai dasar pengambilan keputusan dalam strategi pemasaran, pelayanan, maupun pengelolaan stok barang. Proses ini menjadikan penelitian tidak hanya bersifat teoritis, tetapi juga aplikatif karena langsung berdampak pada efisiensi dan efektivitas pengelolaan toko.

Selain itu, kelebihan lainnya terletak pada penggunaan data aktual dari lingkungan usaha kecil menengah (UKM), yaitu Toko Riski, yang memberikan konteks nyata dan relevan terhadap kebutuhan bisnis lokal. Penelitian ini tidak menggunakan data simulasi atau hipotesis semata, melainkan berangkat dari kebutuhan nyata pelaku usaha dalam memahami pola belanja pelanggan. Hal ini

menunjukkan bahwa penelitian tidak hanya berkontribusi pada pengembangan ilmu pengetahuan di bidang *Data Mining*, tetapi juga memberikan solusi praktis bagi pelaku UMKM yang ingin memanfaatkan teknologi informasi untuk memperkuat daya saingnya. Penggunaan data transaksi asli juga menjadikan hasil analisis lebih akurat dan representatif terhadap situasi yang sebenarnya terjadi di lapangan.

Kelebihan berikutnya adalah penerapan proses penelitian yang sistematis dan terstruktur, dimulai dari tahap pengumpulan data, pembersihan data, transformasi data, hingga evaluasi hasil *Clustering*. Proses ini mengikuti standar *Knowledge Discovery in Databases* (KDD), sehingga menghasilkan analisis yang kuat dan bisa dipertanggungjawabkan secara metodologis. Penelitian ini juga melibatkan konversi atribut kategorikal ke numerikal dan pemanfaatan perangkat lunak Orange, yang membuat proses analisis menjadi lebih efisien dan mudah divisualisasikan. Dengan demikian, penelitian ini tidak hanya memperkaya pemahaman tentang metode *K-Means*, tetapi juga memperkenalkan cara kerja praktis bagi pembaca atau pelaku usaha yang ingin mencoba pendekatan serupa dalam bisnis mereka.