BAB II
LANDASAN TEORI

2.1. Kelapa Sawit

Kelapa sawit (*Elaeis guineensis*) merupakan salah satu komoditas pertanian unggulan di Indonesia yang memiliki peran penting dalam perekonomian nasional. Tanaman ini berasal dari Afrika Barat, namun telah dibudidayakan secara luas di negara-negara tropis, termasuk Indonesia dan Malaysia sebagai dua produsen terbesar dunia. Minyak kelapa sawit yang dihasilkan dari buahnya memiliki berbagai kegunaan, mulai dari bahan makanan seperti minyak goreng dan margarin, hingga bahan baku industri seperti kosmetik, sabun, dan biodiesel. Produktivitasnya yang tinggi dan siklus panen yang terus-menerus menjadikan kelapa sawit sebagai tanaman yang sangat efisien dalam menghasilkan minyak nabati dibandingkan dengan tanaman penghasil minyak lainnya.

Selain manfaat ekonominya, budidaya kelapa sawit juga sering dikaitkan dengan isu lingkungan dan sosial. Pembukaan lahan untuk perkebunan sawit yang tidak terkontrol dapat menyebabkan deforestasi, hilangnya keanekaragaman hayati, dan konflik agraria dengan masyarakat lokal. Oleh karena itu, saat ini mulai dikembangkan sistem budidaya yang lebih berkelanjutan melalui skema sertifikasi seperti RSPO (Roundtable on Sustainable Palm Oil). Di tingkat petani, khususnya petani swadaya, kelapa sawit tetap menjadi sumber pendapatan utama yang dapat meningkatkan kesejahteraan keluarga, terutama di wilayah pedesaan. Dengan pengelolaan yang baik, kelapa sawit berpotensi besar untuk terus memberikan manfaat ekonomi tanpa mengesampingkan aspek sosial dan lingkungan.

2.2. Penjualan

Penjualan merupakan aktivitas inti dalam dunia usaha yang berkaitan langsung dengan pertukaran barang atau jasa untuk mendapatkan keuntungan [3]. Proses ini tidak hanya mencakup transaksi, tetapi juga strategi dalam mengenali kebutuhan pasar, menentukan harga, membangun hubungan dengan pelanggan, dan mengelola distribusi produk [4]. Dalam sektor komoditas seperti kelapa sawit, penjualan memiliki dinamika tersendiri karena dipengaruhi oleh fluktuasi harga pasar, kualitas hasil panen, serta intensitas pasokan dari petani.

Pemanfaatan data penjualan secara analitis dapat membantu pelaku usaha memahami pola transaksi dan perilaku pelanggan. Melalui teknik pengelompokan seperti *K-Means Clustering*, informasi penjualan dapat diklasifikasikan ke dalam kelompok-kelompok berdasarkan kesamaan karakteristik, seperti frekuensi transaksi atau jumlah pembelian [5]. Hal ini memungkinkan pengambilan keputusan yang lebih tepat sasaran, seperti strategi penetapan harga, pelayanan khusus untuk kelompok pelanggan tertentu, atau bahkan pengaturan logistik distribusi sawit agar lebih efisien dan sesuai kebutuhan pasar.

2.3. Penjualan Kelapa Sawit

Penjualan kelapa sawit merupakan salah satu aktivitas ekonomi utama di daerah pedesaan yang memiliki perkebunan sawit. Proses penjualan biasanya dilakukan oleh petani atau pengepul kepada pihak RAM (Rumah Timbang) atau langsung ke pabrik pengolahan kelapa sawit. Dalam proses ini, tandan buah segar (TBS) hasil panen ditimbang untuk menentukan jumlah kilogram yang akan dijual, kemudian dikalikan dengan harga pasar yang berlaku pada saat itu. Harga kelapa

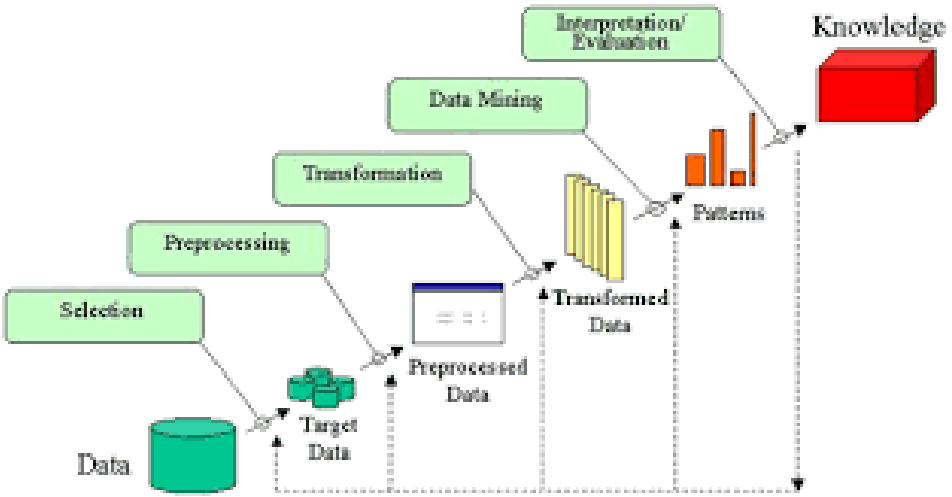
sawit sangat fluktuatif, dipengaruhi oleh faktor-faktor seperti permintaan global, harga minyak nabati dunia, nilai tukar rupiah, serta kebijakan pemerintah. Karena itulah, petani sangat memperhatikan waktu penjualan agar dapat memperoleh keuntungan yang maksimal.

Sistem penjualan kelapa sawit juga semakin berkembang dari waktu ke waktu. Banyak RAM atau tempat penjualan kini telah mengadopsi sistem pencatatan transaksi secara digital guna meningkatkan transparansi dan akurasi data. Selain itu, beberapa tempat penjualan memberikan fasilitas tambahan seperti pinjaman usaha atau sistem tabungan bagi petani yang rutin menjual hasil panennya di tempat tersebut. Adanya hubungan jangka panjang dan saling percaya antara penjual dan pembeli menjadi faktor penting dalam kelancaran aktivitas penjualan sawit. Dengan pengelolaan yang baik dan dukungan teknologi, penjualan kelapa sawit tidak hanya menjadi sumber penghasilan tetapi juga menjadi penggerak ekonomi lokal di banyak wilayah penghasil sawit.

2.4. Knowledge Discovery in Database

Knowledge Discovery in Database (KDD) merupakan proses sistematis dalam menemukan pengetahuan atau pola yang bermanfaat dari sejumlah besar data[6]. Proses ini terdiri dari beberapa tahapan utama, yaitu seleksi data, praproses data, transformasi data, *Data Mining*, dan interpretasi atau evaluasi hasil. Tujuan dari KDD adalah mengubah data mentah menjadi informasi bermakna yang dapat digunakan untuk mendukung pengambilan keputusan [7]. Metode *Data Mining* menjadi inti dari proses KDD, karena pada tahap inilah pola-pola tersembunyi

dalam data diidentifikasi menggunakan teknik statistik, algoritma pembelajaran mesin, maupun kecerdasan buatan [8].



Gambar 2. 1. Knowledge Discovery in Database (KDD)

KDD menjadi kerangka yang tepat dalam proses analisis data penjualan sawit karena mampu menangani data dalam jumlah besar secara sistematis dan terstruktur [9]. Melalui tahapan KDD, data penjualan sawit diolah mulai dari tahap pembersihan dan seleksi hingga proses klasifikasi atau klasterisasi menggunakan algoritma *K-Means* [10]. Hasil akhir dari proses ini dapat memberikan wawasan baru yang sebelumnya tersembunyi dalam data, seperti segmentasi pelanggan atau identifikasi kelompok penjual dengan karakteristik tertentu, yang berguna bagi pengelola RAM dalam pengambilan keputusan bisnis [11].

2.4.1. Data Mining

Data Mining merupakan proses penggalian informasi penting dari sekumpulan data besar yang sebelumnya belum diketahui, dengan tujuan menemukan pola, hubungan, atau tren yang berguna dalam pengambilan keputusan [12] [13]. Teknik ini melibatkan berbagai metode statistik, algoritma kecerdasan

buatan, dan pembelajaran mesin untuk menganalisis data secara mendalam [14] [15]. *Data Mining* banyak diterapkan di berbagai bidang, seperti keuangan, pemasaran, kesehatan, dan industri, karena mampu mengubah data mentah menjadi informasi yang bermakna dan bernilai strategis [16].

Penggunaan *Data Mining* sangat relevan dalam menganalisis data penjualan yang bersifat kompleks dan terus bertambah [17] [18]. Dengan menerapkan algoritma yang tepat, seperti *K-Means Clustering*, data transaksi penjualan sawit dapat dikelompokkan berdasarkan kesamaan karakteristik tertentu [19] [20]. Hasil pengelompokan ini memberikan wawasan baru yang dapat dimanfaatkan untuk mengidentifikasi pola perilaku pelanggan, menyusun strategi pemasaran, serta meningkatkan efisiensi operasional dalam pengelolaan penjualan [21] [22].

2.4.2. *Database* dan *Data Processing*

Database merupakan kumpulan data yang tersimpan secara sistematis dan dapat diakses, dikelola, serta diperbarui dengan mudah menggunakan perangkat lunak tertentu [23]. Dalam dunia digital, pengelolaan data tidak hanya sebatas menyimpan informasi, tetapi juga mencakup proses pemrosesan data seperti pembersihan (cleaning), transformasi, dan integrasi agar data siap dianalisis [24]. Tahapan ini dikenal dengan istilah data *preProcessing* yang sangat penting untuk memastikan bahwa data yang digunakan dalam proses analisis adalah data yang valid, relevan, dan bebas dari noise atau duplikasi yang dapat mengganggu hasil analisis [25] [26].

Pemrosesan data dilakukan untuk menyiapkan data mentah menjadi bentuk yang dapat diterima oleh algoritma analisis, seperti *K-Means Clustering* [27] [28]. Data penjualan sawit yang semula tercatat dalam format transaksi harian perlu disusun ulang dalam format tabel terstruktur, kemudian dinormalisasi agar setiap variabel memiliki skala yang seimbang [29]. Proses ini memungkinkan pengelompokan data berdasarkan kesamaan karakateristik dapat dilakukan secara optimal, sehingga hasil klasterisasi dapat merepresentasikan kondisi riil yang bermanfaat untuk mendukung pengambilan keputusan di lapangan [30].

2.4.3. Visualization

Visualisasi data (*data Visualization*) adalah proses penyajian data dalam bentuk grafis atau visual seperti grafik, diagram, atau peta, dengan tujuan mempermudah pemahaman informasi yang terkandung dalam data [31]. Dengan menggunakan visualisasi, pola, tren, dan anomali dalam data menjadi lebih mudah dikenali dibandingkan jika hanya disajikan dalam bentuk angka atau tabel mentah [32]. Alat visual seperti *scatter plot, bar chart,* dan *pie chart* memungkinkan pengguna melihat hubungan antar variabel serta distribusi data secara lebih intuitif dan informatif [33].

Visualisasi sangat membantu dalam proses analisis *Clustering* karena dapat menunjukkan hasil pengelompokan secara visual dan memperjelas bagaimana objek-objek data terbagi dalam klaster tertentu [34]. Misalnya, melalui scatter plot dua dimensi, setiap klaster dapat ditampilkan dalam warna berbeda sehingga memudahkan identifikasi perbedaan antar kelompok [35]. Hal ini memberikan gambaran konkret tentang distribusi dan kedekatan antar data penjualan sawit, serta

memudahkan pengambilan keputusan berdasarkan hasil klasterisasi yang telah dilakukan menggunakan metode *K-Means*.

2.4.4. Statistik

Statistik merupakan cabang ilmu yang mempelajari cara mengumpulkan, mengolah, menganalisis, menafsirkan, dan menyajikan data agar menghasilkan informasi yang berguna dalam pengambilan keputusan [36]. Statistik terbagi menjadi dua jenis utama, yaitu statistik deskriptif dan statistik inferensial. Statistik deskriptif digunakan untuk menggambarkan atau meringkas data melalui tabel, grafik, rata-rata, median, modus, dan sebagainya [37]. Sementara itu, statistik inferensial digunakan untuk menarik kesimpulan atau membuat prediksi berdasarkan data sampel terhadap populasi yang lebih besar [38].

Penggunaan statistik sangat penting dalam membantu proses analisis data agar lebih terstruktur dan objektif [39]. Dalam pengolahan data penjualan sawit, statistik berperan dalam menyajikan gambaran awal dari data seperti jumlah transaksi, rata-rata volume penjualan, hingga distribusi harga. Informasi statistik ini menjadi dasar penting sebelum dilakukan proses analisis lanjutan seperti penerapan metode *K-Means Clustering*, karena membantu menentukan variabel yang relevan dan pola awal yang mungkin tersembunyi dalam kumpulan data yang ada.

2.4.5. Pattern Recognition

Pattern Recognition adalah cabang ilmu komputer dan kecerdasan buatan yang mempelajari cara mengidentifikasi struktur atau keteraturan di dalam sekumpulan data, lalu mengklasifikasikan atau mengelompokkannya secara

otomatis [40]. Proses ini mencakup tahapan akuisisi data, pra-proses (Normalisasi, reduksi dimensi), pemilihan fitur, penerapan algoritma (seperti *K-Nearest Neighbor*, *Support Vector Machine*, atau *K-Means*), dan evaluasi hasil. Dengan demikian, *Pattern Recognition* memungkinkan sistem menemukan hubungan tersembunyi—baik dalam bentuk pola linier sederhana maupun korelasi kompleks—yang sering luput dari pengamatan manusia, sehingga banyak dimanfaatkan pada visi komputer, pengenalan suara, medis, hingga analisis pasar.

K-Means Clustering sebagai metode *Pattern Recognition* tanpa pengawasan sangat relevan untuk mengekstraksi pola penjualan sawit di RAM BS: data-data seperti kuantitas tandan buah segar, frekuensi transaksi, dan harga jual per kilogram dapat dikelompokkan menjadi klaster-klaster homogen. Hasil klasterisasi ini membantu mengungkap segmentasi pelanggan—misalnya pengepul dengan volume tinggi, penjual musiman, atau petani rutin—yang kemudian bisa dijadikan dasar pengambilan keputusan manajerial, perancangan program loyalitas, dan optimasi rantai pasok.

2.5. Metode *K-Means*

Metode *K-Means Clustering* adalah salah satu teknik dalam *Data Mining* yang digunakan untuk melakukan pengelompokan data ke dalam sejumlah klaster berdasarkan tingkat kemiripan data [41]. Algoritma ini bersifat unsupervised learning, artinya tidak memerlukan label atau target output saat proses pelatihan data dilakukan. Prinsip kerja dari *K-Means* adalah dengan menentukan sejumlah klaster awal (*k*), lalu secara iteratif menempatkan data ke dalam klaster berdasarkan jarak terdekat terhadap titik pusat (*Centroid*) masing-masing klaster [42]. Setelah

itu, *Centroid* akan dihitung ulang berdasarkan rata-rata dari semua data dalam klaster tersebut, dan proses ini diulang hingga tidak terjadi lagi perubahan signifikan dalam pembagian klaster atau jumlah iterasi telah tercapai [43]. Kelebihan metode ini adalah kesederhanaannya dalam implementasi dan kecepatannya dalam menangani jumlah data yang besar, meskipun ada kelemahan seperti kepekaan terhadap nilai *k* yang ditentukan dan data *outlier* [44].

Penerapan metode ini sangat cocok dalam mengelompokkan data numerik yang memiliki atribut serupa, misalnya volume penjualan, frekuensi transaksi, atau nilai rata-rata pembelian. Dalam kasus penjualan sawit, metode ini dapat digunakan untuk mengelompokkan data transaksi berdasarkan kesamaan perilaku penjual, sehingga menghasilkan klaster yang merepresentasikan tipe-tipe pelanggan atau pola transaksi yang berbeda. Dengan proses klasterisasi ini, pengelola dapat memahami kelompok penjual mana yang melakukan transaksi dalam jumlah besar dan konsisten, kelompok yang fluktuatif, atau yang hanya melakukan transaksi dalam jumlah kecil. Hasil klasterisasi dapat digunakan sebagai dasar dalam merancang strategi layanan, menentukan prioritas pembinaan, bahkan pengambilan keputusan operasional secara lebih terarah dan berbasis data.

Dalam proses pelaksanaannya, algoritma *K-Means* memerlukan beberapa tahap, di antaranya adalah pra-pemrosesan data (*preProcessing*), penentuan jumlah klaster (*k*), pemilihan *Centroid* awal, proses iteratif pembentukan klaster, dan evaluasi hasil klasterisasi. Pra-pemrosesan biasanya dilakukan dengan normalisasi data agar skala antar variabel seragam. Penentuan nilai *k* dapat dilakukan dengan teknik seperti *Elbow Method* atau *Silhouette Score* untuk mendapatkan jumlah

klaster yang paling optimal. Setelah itu, proses klasterisasi dilakukan hingga stabil. Evaluasi hasil klaster biasanya dilakukan untuk memastikan bahwa setiap klaster memiliki homogenitas internal yang tinggi dan heterogenitas antar-klaster yang baik.

Secara keseluruhan, *K-Means* bukan hanya mampu mengelompokkan data menjadi beberapa segmen yang jelas, tetapi juga mampu membuka wawasan baru terhadap karakteristik tersembunyi dalam data yang sebelumnya tidak terlihat. Dalam praktiknya, metode ini banyak digunakan di berbagai bidang seperti pemasaran, perbankan, retail, dan analisis perilaku konsumen. Fleksibilitas dan kemudahan implementasi menjadi keunggulan utama dari *K-Means*, terutama jika diterapkan pada data transaksional atau bisnis yang memiliki pola-pola numerik yang dapat dikelompokkan secara statistik.

2.5.1. Fungsi Metode *K-Means*

Metode *K-Means Clustering* berfungsi sebagai teknik pengelompokan data yang bertujuan untuk membagi sekumpulan data ke dalam beberapa kelompok (klaster) berdasarkan kemiripan karakteristik tertentu [45]. Algoritma ini bekerja dengan menentukan pusat klaster (*Centroid*), kemudian mengelompokkan data berdasarkan jarak terdekat ke pusat tersebut. Proses ini berlangsung secara iteratif sampai posisi pusat klaster stabil dan tidak lagi berubah [46]. Dengan cara ini, data yang sebelumnya tidak terstruktur dapat dibagi menjadi kelompok-kelompok yang lebih mudah dianalisis.

Metode ini sangat membantu dalam proses identifikasi pola tersembunyi dari data yang beragam. Misalnya, pada data penjualan sawit, metode *K-Means* dapat digunakan untuk mengelompokkan transaksi berdasarkan volume pembelian, harga jual, atau frekuensi transaksi dari masing-masing pelanggan atau pengepul. Dengan terbentuknya kelompok-kelompok yang seragam, pihak pengelola penjualan dapat mengetahui karakteristik dari setiap klaster, seperti klaster pelanggan besar, pelanggan sedang, dan pelanggan kecil. Informasi ini sangat berguna dalam menentukan strategi pelayanan, pemasaran, atau pengambilan keputusan manajerial lainnya.

Fungsi lain dari *K-Means* adalah menyederhanakan proses visualisasi dan analisis data dalam jumlah besar. Data yang awalnya tersebar tanpa struktur menjadi lebih terorganisir dalam beberapa kelompok yang memiliki makna tertentu. Dengan adanya klaster, analisis lebih lanjut dapat difokuskan pada tiap kelompok secara spesifik, sehingga menghasilkan wawasan yang lebih tajam dan relevan. Hal ini menjadikan *K-Means* sebagai salah satu metode yang efektif untuk mendukung pengambilan keputusan berbasis data.

2.5.2. Ciri-Ciri Metode *K-Means*

Metode *K-Means Clustering* memiliki ciri utama yaitu bersifat unsupervised learning, artinya tidak memerlukan label atau target output dalam proses pengelompokan data. Algoritma ini bekerja berdasarkan kemiripan antar data, di mana objek-objek yang memiliki karakteristik serupa akan dikelompokkan ke dalam klaster yang sama. Penentuan klaster dilakukan berdasarkan nilai *Centroid* (titik pusat) yang dihitung secara iteratif hingga mencapai kondisi optimal atau

konvergen. Inilah yang menjadikan metode ini cocok untuk menganalisis data tanpa informasi awal mengenai kelompok-kelompoknya.

Ciri lainnya dari *K-Means* adalah kemampuannya dalam bekerja dengan data numerik dan terukur, karena perhitungan jarak antar data sangat penting dalam menentukan keanggotaan klaster. Umumnya digunakan *Euclidean Distance* sebagai ukuran jarak antar objek. Prosesnya melibatkan pembagian data ke dalam sejumlah klaster yang telah ditentukan sebelumnya, sehingga pengguna harus menentukan nilai *k* (jumlah klaster) di awal proses. Meskipun terlihat sederhana, pemilihan nilai *k* yang tepat sangat krusial karena dapat memengaruhi akurasi dan makna dari hasil klasterisasi.

Selain itu, metode ini juga memiliki karakteristik berupa kecepatan dan efisiensi yang tinggi, terutama saat digunakan pada dataset yang besar namun bersifat homogen. Namun, metode ini cenderung sensitif terhadap outlier dan nilai ekstrem, serta dapat menghasilkan hasil yang berbeda jika inisialisasi *Centroid* awal berbeda. Oleh karena itu, seringkali dilakukan beberapa kali iterasi atau pengujian untuk memastikan kestabilan hasil. Dengan karakteristik-karakteristik tersebut, *K-Means* menjadi salah satu metode *Clustering* yang paling banyak digunakan dalam analisis data praktis, termasuk dalam pengelompokan transaksi penjualan seperti di penelitian ini.

2.5.3. Langkah-Langkah Metode *K-Means*

Langkah pertama dalam metode *K-Means* adalah menentukan jumlah klaster (k) yang ingin dibentuk. Pemilihan nilai *k* ini sangat penting karena akan

memengaruhi kualitas hasil klasterisasi. Nilai *k* dapat ditentukan berdasarkan pengetahuan domain, pengamatan visual terhadap data, atau dengan bantuan metode seperti *Elbow Method*, yang digunakan untuk menemukan titik optimal dari jumlah klaster dengan membandingkan nilai *within-Cluster sum of squares* (WCSS).

Langkah berikutnya adalah inisialisasi *Centroid* secara acak sesuai jumlah klaster yang telah ditentukan. Setelah *Centroid* awal ditentukan, setiap data dalam dataset akan dihitung jaraknya ke masing-masing *Centroid* menggunakan rumus tertentu, umumnya Euclidean Distance. Data kemudian dikelompokkan ke dalam klaster berdasarkan *Centroid* terdekat. Setelah semua data terbagi ke dalam klaster, sistem akan menghitung ulang posisi *Centroid* baru dari masing-masing klaster berdasarkan rata-rata posisi semua data dalam klaster tersebut.

Langkah-langkah di atas akan diulang secara iteratif hingga posisi *Centroid* tidak lagi berubah secara signifikan atau jumlah iterasi maksimum tercapai. Proses ini disebut konvergensi, dan ketika tercapai, maka hasil klasterisasi dianggap final. Hasil akhir dari metode *K-Means* adalah pembagian data ke dalam klaster-klaster yang memiliki karakteristik mirip, yang dapat digunakan untuk analisis lanjutan seperti segmentasi pelanggan, pola pembelian, atau pengambilan keputusan strategis lainnya.

2.6. Model *Cluster*

Model *Cluster* atau klasterisasi merupakan salah satu metode dalam *Data Mining* yang digunakan untuk mengelompokkan data ke dalam sejumlah kelompok

atau klaster berdasarkan kemiripan atau kedekatan antar data. Dalam teknik ini, data yang memiliki karakteristik serupa akan ditempatkan dalam satu kelompok, sedangkan data yang berbeda akan masuk ke kelompok lainnya. Tujuan utama dari pembentukan klaster adalah untuk menemukan struktur alami dalam data, sehingga pola-pola yang tersembunyi dapat terungkap tanpa perlu informasi label sebelumnya (unsupervised learning). Salah satu algoritma paling populer yang digunakan dalam teknik klasterisasi adalah *K-Means Clustering*, yang bekerja dengan mempartisi data ke dalam k kelompok berdasarkan jarak terdekat ke pusat klaster (*Centroid*). Proses ini berlangsung secara iteratif hingga komposisi klaster menjadi stabil. Klasterisasi sangat berguna dalam berbagai bidang, seperti pemasaran, pengelompokan pelanggan, segmentasi citra, hingga analisis perilaku pengguna.

Penggunaan model *Cluster* sangat relevan untuk mengungkap pola tersembunyi dalam data penjualan yang sebelumnya belum dianalisis secara sistematis. Dalam konteks transaksi penjualan sawit, klasterisasi memungkinkan untuk mengelompokkan data berdasarkan volume penjualan, frekuensi transaksi, atau variasi harga. Dengan pendekatan ini, dapat diketahui kelompok penjual yang rutin bertransaksi dalam jumlah besar, kelompok menengah, serta kelompok kecil yang mungkin bersifat musiman. Masing-masing klaster tersebut tentu memiliki karakteristik yang berbeda dan dapat dimanfaatkan oleh pengelola usaha untuk menyusun strategi pelayanan, penentuan harga, hingga pendekatan relasional terhadap pelanggan. Model *Cluster* memberikan gambaran menyeluruh tanpa harus

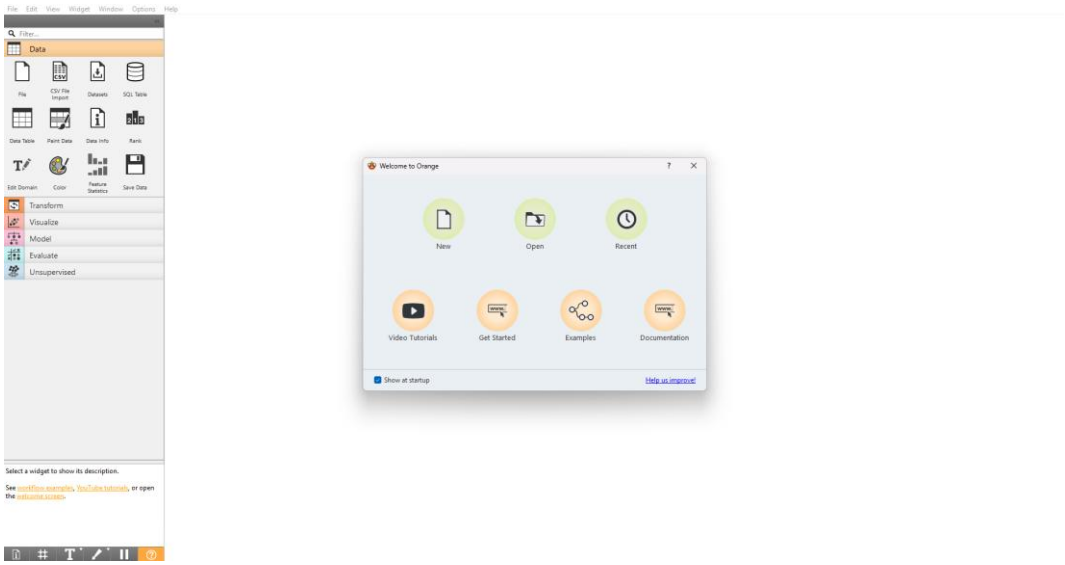
memulai dari asumsi tertentu, sehingga cocok untuk analisis data yang masih bersifat mentah dan tidak terstruktur.

Lebih lanjut, penerapan *K-Means Clustering* dalam membentuk model *Cluster* biasanya diawali dengan menentukan jumlah klaster (k) yang diinginkan. Penentuan jumlah klaster ini dapat dilakukan dengan beberapa pendekatan, seperti metode Elbow atau Silhouette. Setelah k ditentukan, algoritma akan membagi data ke dalam klaster berdasarkan nilai *Centroid* yang diperbarui secara terus-menerus. Setiap iterasi akan memperkecil selisih jarak antara data dan pusat klasternya hingga mencapai konvergensi. Hasil akhirnya adalah pemetaan data ke dalam beberapa klaster yang memiliki pola dan kecenderungan yang jelas. Hasil tersebut tidak hanya dapat divisualisasikan untuk keperluan analisis, tetapi juga bisa menjadi dasar untuk pengambilan keputusan strategis dalam konteks operasional dan bisnis.

Model *Cluster* bukan hanya alat teknis, melainkan juga pendekatan strategis dalam memahami struktur data yang kompleks. Dengan menyederhanakan data ke dalam kelompok-kelompok yang memiliki makna, pengambil keputusan dapat lebih mudah mengenali segmen potensial, mendeteksi anomali, serta merumuskan kebijakan berbasis bukti. Dalam praktiknya, model *Cluster* juga sering dipadukan dengan visualisasi data, seperti grafik sebar (scatter plot) atau diagram radar, untuk membantu memahami distribusi antar klaster secara intuitif. Kemampuan model ini dalam menyajikan informasi secara ringkas dan terfokus menjadikannya sangat berguna dalam berbagai konteks analisis, termasuk untuk sektor agribisnis seperti penjualan hasil panen kelapa sawit.

2.7. Aplikasi Orange

Aplikasi Orange adalah salah satu perangkat lunak open-source yang dirancang khusus untuk kebutuhan analisis data dan pembelajaran mesin (machine learning). Aplikasi ini memiliki antarmuka visual berbasis drag and drop yang sangat ramah bagi pengguna, bahkan bagi pemula yang belum mahir dalam pemrograman. Orange mendukung berbagai teknik analisis data, seperti klasifikasi, klasterisasi, regresi, asosiasi, dan visualisasi data. Keunggulan utamanya terletak pada fleksibilitas dan kemudahan integrasi antar komponen melalui widgets yang dapat disusun sesuai kebutuhan proses analisis. Dengan berbagai fitur ini, Orange telah banyak digunakan dalam penelitian dan pembelajaran karena mampu menyederhanakan proses eksplorasi dan modeling data tanpa perlu menulis kode secara manual.



Gambar 2. 2. Tampilan Awal Aplikasi Orange

Penggunaan Orange sangat relevan untuk mendukung proses pengolahan data penjualan sawit karena mampu menyajikan alur kerja analisis yang sistematis dan interaktif. Melalui Orange, data dapat diimpor, dibersihkan, dan langsung dianalisis

dengan metode *K-Means Clustering* hanya dengan menghubungkan beberapa *widget* seperti File, *Data Table*, *Preprocess, K-Means*, dan *Scatter Plot*. Hal ini memungkinkan pengguna untuk dengan cepat melihat hasil klasterisasi dan mengevaluasi pola-pola yang muncul dari data penjualan. Selain itu, Orange juga menyediakan visualisasi data dalam bentuk grafik yang interaktif, sehingga memudahkan interpretasi terhadap hasil pengelompokan yang dihasilkan oleh algoritma yang digunakan.

Kelebihan lainnya, Orange juga menyediakan fitur evaluasi model dan penyimpanan hasil analisis yang dapat digunakan kembali atau dibandingkan dengan metode lain. Dengan alur kerja yang jelas dan minim kesalahan input, Orange menjadi alat bantu yang sangat efektif dalam menerapkan metode *Data Mining* secara efisien. Kombinasi antara kemudahan penggunaan, kemampuan analitis, dan visualisasi interaktif menjadikan Orange sebagai solusi praktis dalam mendukung kegiatan penelitian berbasis data, termasuk pada sektor pertanian dan perdagangan hasil kebun seperti penjualan kelapa sawit.

2.8. Langkah-Langkah Penerapan Aplikasi Orange

Dalam penelitian ini, penerapan aplikasi Orange *Data Mining* dilakukan sebagai alat bantu untuk proses klasifikasi dan analisis data yang berkaitan dengan objek penelitian. Orange merupakan aplikasi open-source berbasis visualisasi yang dirancang untuk mendukung proses analisis data tanpa harus menulis kode pemrograman secara langsung. Proses penggunaannya dilakukan secara bertahap mulai dari proses impor data, pembersihan data, eksplorasi, pemodelan, hingga evaluasi model. Langkah pertama yang dilakukan adalah menyiapkan dataset dalam

format yang dapat dikenali oleh Orange, seperti file CSV atau Excel. Data ini kemudian dimasukkan ke dalam canvas kerja Orange menggunakan widget File atau Datasets.

Setelah data berhasil dimuat ke dalam sistem, tahap selanjutnya adalah melakukan data preprocessing. Pada tahap ini, widget seperti Select Columns, Edit Domain, dan Data Table digunakan untuk mengatur variabel-variabel yang akan dianalisis, baik sebagai atribut input maupun target klasifikasi. Data yang tidak relevan atau mengandung nilai kosong biasanya akan dibersihkan atau dikelola terlebih dahulu. Proses ini sangat penting agar model yang dibangun nantinya memiliki kualitas data yang baik. Selain itu, pengguna juga dapat menerapkan normalisasi atau pengelompokan nilai (binning) pada data numerik agar lebih sesuai dengan algoritma klasifikasi yang digunakan.

Tahap ketiga adalah eksplorasi data, di mana peneliti dapat menggunakan berbagai widget visualisasi seperti *Box Plot*, Distributions, Scatter Plot, dan Data Info untuk memahami distribusi data, hubungan antar atribut, serta potensi outlier. Tahapan eksplorasi ini penting untuk memperoleh pemahaman awal tentang karakteristik dataset dan menentukan strategi analisis yang tepat. Dengan bantuan visualisasi, pengguna dapat melihat pola-pola tertentu dalam data yang mungkin tidak terlihat hanya dengan membaca angka mentah. Informasi ini akan sangat berguna saat membangun dan menyesuaikan model klasifikasi.

Selanjutnya adalah tahap pemodelan, di mana pengguna dapat memilih berbagai algoritma yang tersedia dalam Orange, seperti *Naïve Bayes, k-Nearest*

Neighbor (kNN), Decision Tree, atau Support Vector Machine (SVM). Algoritma tersebut dapat dihubungkan langsung dengan data yang telah diproses melalui widget seperti Test & Score dan Confusion Matrix untuk mengukur performa model. Widget Train Test Split atau Cross Validation biasanya digunakan untuk membagi data menjadi data latih dan data uji agar hasil evaluasi lebih objektif. Hasil dari evaluasi ini meliputi akurasi, presisi, recall, F1-score, dan AUC, yang dapat dibandingkan antar model untuk memilih yang terbaik.

Langkah terakhir adalah interpretasi dan pengambilan keputusan berdasarkan output yang dihasilkan. Orange menyediakan visualisasi dari hasil klasifikasi, seperti grafik *confusion matrix* dan *ROC Curve*, yang membantu peneliti dalam memahami seberapa baik model dapat memprediksi data baru. Jika diperlukan, model dapat dioptimasi ulang dengan mengganti parameter atau melakukan pengolahan data tambahan. Hasil akhir dari seluruh proses ini digunakan untuk menjawab rumusan masalah dalam penelitian dan memberikan rekomendasi atau kesimpulan yang berbasis data. Dengan alur kerja yang intuitif dan visual, Orange memudahkan peneliti dari berbagai latar belakang untuk melakukan analisis data secara efektif dan efisien.

2.9. Kelebihan dan Kekurangan Metode *K-Means*

Metode *K-Means Clustering* merupakan salah satu algoritma unsupervised learning yang paling populer dalam proses pengelompokan data atau *Clustering*. Prinsip kerja dari metode ini adalah membagi sekumpulan data ke dalam sejumlah klaster yang telah ditentukan sebelumnya (jumlah *k*), berdasarkan kedekatan nilai fitur antar data. Proses ini dilakukan dengan menghitung *Centroid* atau pusat dari

setiap klaster, lalu mengelompokkan data berdasarkan jarak terdekat ke *Centroid* tersebut. Kelebihan utama dari *K-Means* adalah kemampuannya untuk bekerja dengan sangat efisien terhadap dataset yang besar dan kompleks, serta hasil visualisasi yang mudah dipahami. Selain itu, algoritma ini memiliki proses komputasi yang relatif cepat karena hanya memerlukan perhitungan sederhana untuk menghitung jarak antar titik data dan pusat klaster.

Kelebihan ini sangat relevan ketika diterapkan dalam analisis data penjualan sawit di RAM BS, di mana jumlah transaksi yang terus bertambah membutuhkan metode yang efisien dalam waktu dan komputasi. Melalui pendekatan *K-Means*, data transaksi dapat dikelompokkan menjadi beberapa segmen, misalnya berdasarkan volume penjualan atau pola pembelian. Hal ini memungkinkan pihak pengelola untuk mengidentifikasi kelompok pelanggan tertentu, seperti pelanggan aktif, pelanggan musiman, atau pelanggan dengan jumlah pembelian kecil namun konsisten. Dengan segmentasi tersebut, strategi layanan dapat disesuaikan secara lebih efektif dan tepat sasaran, misalnya dengan memberikan insentif kepada kelompok tertentu atau memperkuat hubungan dengan pelanggan yang paling menguntungkan.

Meski demikian, metode *K-Means* juga memiliki sejumlah kekurangan yang perlu diperhatikan dalam penerapannya. Salah satu kelemahan utamanya adalah kepekaan terhadap pemilihan jumlah klaster (k) di awal proses. Jika nilai k yang dipilih tidak sesuai dengan karakteristik data, maka hasil klasterisasi bisa jadi tidak akurat atau kurang representatif. Selain itu, *K-Means* juga sangat dipengaruhi oleh posisi awal *Centroid*. Pemilihan *Centroid* yang tidak tepat pada iterasi pertama

dapat menyebabkan hasil akhir cenderung konvergen ke solusi lokal, bukan solusi global terbaik. Dalam beberapa kasus, hal ini membuat algoritma perlu dijalankan beberapa kali dengan inisialisasi berbeda untuk mendapatkan hasil yang stabil.

Keterbatasan lainnya terletak pada asumsi bahwa klaster berbentuk bulat atau memiliki distribusi data yang seimbang. Jika data memiliki bentuk distribusi yang tidak teratur atau mengandung outlier, maka *K-Means* bisa salah dalam mengelompokkan data. Hal ini tentu dapat mempengaruhi kualitas hasil analisis, terutama jika data penjualan yang digunakan memiliki variasi ekstrim antar transaksi. Oleh karena itu, sebelum menerapkan metode ini, penting untuk melakukan proses normalisasi dan pembersihan data, serta melakukan analisis eksploratif terlebih dahulu agar hasil klasterisasi menjadi lebih bermakna dan relevan.

Pada pendekatan tambahan seperti penggunaan metode *Elbow* untuk menentukan nilai *k* yang optimal, serta inisialisasi *Centroid* menggunakan metode *K-Means++*, dapat diterapkan. Selain itu, *K-Means* sebaiknya digunakan sebagai bagian dari pendekatan analisis yang lebih luas, seperti dikombinasikan dengan visualisasi data dan validasi klaster menggunakan indeks seperti *Silhouette Score*. Dengan memperhatikan kelebihan dan kekurangannya secara bijak, *K-Means* tetap menjadi alat yang kuat dan praktis dalam membantu pengambilan keputusan berbasis data, terutama di sektor-sektor yang mengandalkan segmentasi pelanggan atau transaksi seperti pada sistem penjualan hasil pertanian.

2.10. Kelebihan Penelitian

Metode *Data Mining* telah menjadi pendekatan yang sangat berguna dalam mengolah data berjumlah besar, terutama ketika informasi yang tersembunyi di dalamnya tidak mudah ditemukan secara manual. Salah satu algoritma yang cukup populer adalah *K-Means Clustering*, yang dikenal efektif dalam mengelompokkan data ke dalam beberapa klaster berdasarkan kemiripan karakteristik. Kelebihan utama dari metode ini adalah kesederhanaannya dalam proses perhitungan, kecepatan dalam memproses data dalam jumlah besar, serta fleksibilitasnya dalam diterapkan di berbagai bidang, mulai dari pemasaran, keuangan, hingga sektor pertanian. Dengan adanya pengelompokan data, hasil analisis dapat memberikan gambaran yang lebih tajam dan terarah, yang pada akhirnya sangat membantu dalam pengambilan keputusan yang berbasis data.

Penggunaan metode *K-Means Clustering* memberikan keunggulan tersendiri karena dapat mengelompokkan transaksi penjualan sawit ke dalam kelompok-kelompok dengan karakteristik yang mirip, misalnya berdasarkan jumlah pembelian, frekuensi transaksi, atau variasi harga. Hasil dari klasterisasi ini dapat menjadi dasar untuk memahami perilaku pelanggan, mengidentifikasi pengepul aktif dan pasif, hingga membantu menyusun strategi pelayanan yang lebih tepat sasaran. Pendekatan ini tidak hanya meningkatkan efisiensi pengelolaan data, tetapi juga membuka peluang bagi pelaku usaha untuk lebih mengenal segmen pasar mereka. Hal ini tentu menjadi nilai tambah dibandingkan pencatatan transaksi konvensional yang selama ini hanya bersifat administratif.

Selain itu, penelitian ini juga memiliki kelebihan dari sisi kebermanfaatan langsung yang ditawarkan. Hasil dari analisis klaster dapat diterapkan secara nyata dalam perencanaan operasional dan strategi jangka panjang oleh pengelola RAM. Informasi yang diperoleh tidak hanya bersifat deskriptif, melainkan juga aplikatif karena dapat menjadi dasar untuk tindakan seperti penyesuaian harga, peningkatan layanan bagi kelompok pelanggan tertentu, atau identifikasi potensi loyalitas pelanggan. Dengan kata lain, penelitian ini mampu menjembatani antara teknologi pengolahan data modern dengan kebutuhan nyata di lapangan, khususnya dalam pengelolaan usaha lokal berbasis hasil pertanian seperti kelapa sawit.

2.11. Penelitian Terdahulu

Tabel 2. 1. Penelitian Terdahulu

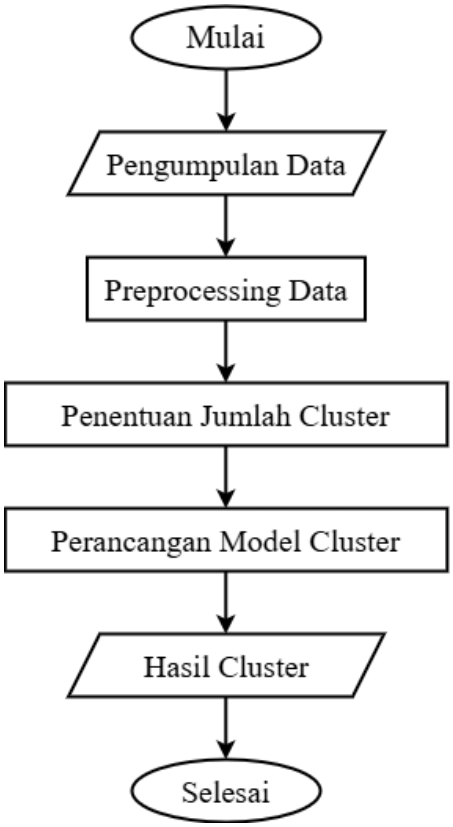
Referensi Penelitian	1
Judul	KLASTERISASI DATA PENJUALAN BERDASARKAN WILAYAH MENGGUNAKAN METODE *K-MEANS* PADA PT XYZ
Nama	1Elin Mayoana Fitri, 2,*)Ryan Randy Suryono & 3Agus Wantoro
Tahun	2023
Hasil	PT XYZ, perusahaan distribusi minuman ringan, mengalami kesulitan dalam mengelola data transaksi penjualan cabang secara manual. Dengan menggunakan algoritma *K-Means Clustering* dan bahasa pemrograman Python, penelitian ini berhasil mengklasifikasikan penjualan ke dalam tiga klaster, memberikan rekomendasi untuk strategi penjualan dan meningkatkan keuntungan Perusahaan [1].
Referensi Penelitian	2

Judul	Implementasi Algoritma *K-Means* Clustring dalam Penetuan Siswa Kelas Unggulan
Nama	Ari Sulistiyawati 1,*, Eko Supriyanto2
Tahun	2023
Hasil	Penelitian ini bertujuan mengembangkan sistem informasi berbasis web dengan algoritma *K-Means Clustering* untuk mengelola data penilaian terpusat, memfasilitasi pembentukan kelas unggulan berdasarkan kemampuan siswa. Implementasi pada satu sekolah di Lampung menghasilkan aplikasi yang efektif, memperoleh dua klaster setiap kelas, dan mendukung pengambilan keputusan dalam menyusun kelas unggulan dengan total 192 siswa. Evaluasi pengguna menggunakan Technology Acceptance Model (TAM) menunjukkan bahwa sistem informasi ini dianggap mudah digunakan dan bermanfaat untuk mengelompokkan siswa kelas unggulan [2].
Referensi Penelitian	3
Judul	META-HEURISTIC ALGORITHMS FOR *K-MEANS CLUSTERING*: A REVIEW
Nama	Alan Fuad Jahwar, Adnan Mohsin Abdulazeez
Tahun	2021
Hasil	Peningkatan ketersediaan data telah memicu minat pada pendekatan pengelompokan, dan algoritma Meta-Heuristic dianggap lebih efektif daripada algoritma optimasi standar dalam menangani masalah optimasi yang sebelumnya dianggap kelemahan dalam algoritma *K-Means*. Dalam tulisan ini, dilakukan tinjauan terhadap algoritma *Clustering K-Means* dan algoritma meta-heuristik untuk menggambarkan pendekatan yang lebih canggih dalam mengelola dan mengidentifikasi pola data besar [3].
Referensi Penelitian	4

Judul	A Novel *K-Means Clustering* Algorithm with a Noise Algorithm for Capturing Urban Hotspots
Nama	Xiaojuan Ran 1,2, Xiangbing Zhou 2,3,*, Mu Lei 4 , Worawit Tepsan 1 and Wu Deng 2,5,*
Tahun	2021
Hasil	Dalam menghadapi masalah kemacetan perkotaan, algoritma pengelompokan *K-Means* telah diakui sebagai pendekatan efektif dalam perencanaan jalan. Namun, sulit menentukan jumlah klaster dan menginisialisasi pusat klaster dengan sensitif. Untuk mengatasi masalah ini, sebuah algoritma *K-Means* baru berdasarkan algoritma kebisingan dikembangkan, menggunakan evaluasi indeks seperti DB, PBM, SC, dan SSE untuk menghasilkan pengelompokan yang lebih baik dengan memverifikasi efektivitasnya menggunakan lima kumpulan data GPS taksi dari berbagai kota [4].

2.12. Kerangka Kerja Penelitian

Kerangka kerja penelitian ini disusun untuk memberikan alur sistematis dalam proses analisis data penjualan sawit menggunakan metode *K-Means Clustering*. Kerangka ini mencakup serangkaian tahapan yang saling berkaitan, mulai dari pengumpulan data hingga interpretasi hasil klasterisasi. Tujuan utamanya adalah mengelompokkan data transaksi penjualan sawit berdasarkan kemiripan karakteristik tertentu, sehingga dapat memberikan informasi yang berguna dalam pengambilan keputusan strategis. Setiap tahap dirancang agar analisis dapat berjalan efisien, akurat, dan relevan terhadap kebutuhan manajerial di RAM BS.



Gambar 2. 3. Kerangka Kerja Penelitian

Untuk mencapai hasil analisis yang optimal dalam proses klasterisasi data penjualan sawit, diperlukan langkah-langkah sistematis dan terstruktur. Berikut ini merupakan tahapan-tahapan yang dilakukan dalam penerapan metode *K-Means*, mulai dari pengumpulan data hingga diperolehnya hasil klaster yang siap dianalisis lebih lanjut.

1. Pengumpulan Data: Pada tahap ini dilakukan pengambilan data penjualan sawit dari sumber yang tersedia di RAM BS. Data ini dapat berupa catatan transaksi, jumlah pembelian, harga, dan identitas pelanggan.
2. Pre*Processing* Data: Data yang telah dikumpulkan dibersihkan dari duplikasi, data kosong, dan nilai ekstrem yang dapat mengganggu proses analisis.

Selanjutnya dilakukan transformasi atau normalisasi agar data berada dalam skala yang sesuai untuk proses klasterisasi.

3. Penentuan Jumlah *Cluster*: Peneliti menentukan jumlah klaster (nilai k) yang optimal dengan menggunakan metode tertentu, seperti Elbow Method. Penentuan jumlah klaster sangat penting karena akan mempengaruhi hasil segmentasi data.
4. Perancangan Model *Cluster*: Pada tahap ini, algoritma *K-Means* dijalankan dengan memasukkan data dan jumlah klaster yang telah ditentukan. Proses ini melibatkan perhitungan jarak antar data dan penentuan pusat klaster (*Centroid*) secara iteratif.
5. Hasil *Cluster*: Hasil dari proses klasterisasi menampilkan kelompok-kelompok data berdasarkan kemiripan karakteristik. Setiap klaster menggambarkan pola tertentu dalam data penjualan yang dapat dianalisis lebih lanjut.