

BAB II

LANDASAN TEORI

2.1 Konsep Dasar Produksi

Produksi merupakan suatu proses kegiatan yang dilakukan untuk mengubah berbagai input atau sumber daya menjadi output berupa barang dan jasa yang memiliki nilai guna lebih tinggi. Dalam perspektif ekonomi, produksi dipahami sebagai aktivitas penciptaan nilai tambah melalui pengolahan faktor-faktor produksi seperti bahan baku, tenaga kerja, modal, dan teknologi. Produksi juga merupakan serangkaian aktivitas yang mentransformasikan input menjadi output demi memenuhi kebutuhan konsumen, yang juga menunjuk pada pentingnya efisiensi dalam proses yang dijalankan (Ahnar-Winarno et al., 2021).

Dalam konteks manajemen operasional, produksi dianggap sebagai sistem terintegrasi. Keberhasilan sistem produksi bergantung pada pengelolaan yang sistematis untuk mencapai keseimbangan antara kapasitas, permintaan, dan sumber daya yang tersedia. Oleh karena itu, ketepatan perencanaan produksi menjadi sangat penting, terutama pada industri dengan permintaan yang fluktuatif, di mana ketidakakuratan dapat mengakibatkan berbagai permasalahan dalam rantai pasok.

2.1.1 Produksi Tempe

Produksi tempe merupakan contoh produksi pangan tradisional berbasis kedelai yang melalui proses fermentasi. Proses ini meliputi pemilihan dan pencucian kedelai, perendaman, perebusan, inokulasi ragi, fermentasi, dan

pengemasan. Setiap tahapan tersebut sangat berpengaruh terhadap kualitas dan kuantitas tempe yang dihasilkan. Menurut studi, permintaan pasar dan ketersediaan bahan baku, seperti kedelai dan ragi, adalah dua faktor utama yang memengaruhi produksi tempe. Ketidakseimbangan antara permintaan dan kapasitas produksi dapat berdampak pada efisiensi operasional dan keberlanjutan usaha (Abdilah et al., 2023).

Bagi UMKM tempe, perencanaan produksi yang akurat sangat penting untuk penyesuaian terhadap fluktuasi permintaan. Pendekatan berbasis data dan prediktif dapat digunakan untuk merencanakan volume produksi secara lebih sistematis, meningkatkan efisiensi penggunaan sumber daya, dan mengurangi risiko pemborosan (Jannah & Lusiani, 2023).

2.2 Data Penjualan

Data penjualan adalah kumpulan informasi yang mencatat aktivitas transaksi penjualan suatu produk dalam periode tertentu. Informasi ini meliputi jumlah produk yang terjual, waktu penjualan, frekuensi transaksi, serta pola permintaan konsumen. Bagi UMKM tempe, data penjualan menjadi rujukan krusial untuk menentukan volume produksi yang sesuai dengan kebutuhan aktual, menghindari *overproduction* atau *underproduction* yang dapat merugikan bisnis. Dengan analisis data penjualan, produsen dapat mengidentifikasi pola musiman, tren konsumen, dan respons terhadap faktor eksternal seperti harga atau promosi, sehingga perencanaan produksi menjadi lebih efisien dan responsif. Data penjualan dapat digunakan untuk memahami perilaku konsumen dan sebagai dasar pengambilan keputusan strategis, terutama dalam perencanaan produksi dan

distribusi (Prastiwi et al., 2021). Selain itu, data penjualan memfasilitasi identifikasi tren dan pola musiman, yang penting untuk analisis lebih lanjut. Pemanfaatan data penjualan historis yang tepat dapat meningkatkan akurasi peramalan permintaan dan mendukung efisiensi operasional perusahaan (Suryani et al., 2025).

2.2.1 Data Penjualan dalam Perencanaan Produksi

Perencanaan produksi melibatkan penentuan jumlah produk yang harus diproduksi dalam periode tertentu untuk memenuhi permintaan pasar secara optimal. Data penjualan historis memainkan peran sentral dalam estimasi kebutuhan produksi, memungkinkan produsen untuk lebih efisien dalam menyesuaikan kapasitas produksi (Suryani et al., 2025). Dalam industri pangan, ketepatan perencanaan produksi menjadi sangat penting karena produk tersebut sering kali mudah rusak. Kesalahan dalam memperkirakan permintaan dapat menyebabkan pemborosan atau kehilangan peluang penjualan. (Angelia et al., 2025) mencatat bahwa integrasi data penjualan dalam perencanaan produksi dapat mengurangi risiko ketidakseimbangan pasokan dan permintaan, sekaligus meningkatkan kinerja rantai pasok. Selain itu, usaha kecil dan menengah (UMKM) yang menggunakan data penjualan untuk perencanaan produksi dapat merespons dinamika pasar dengan lebih baik, menunjukkan pentingnya data penjualan dalam pengambilan keputusan yang rasional (Prastiwi et al., 2021).

2.2.2 Peran Data Penjualan terhadap Volume Produksi

Volume produksi merupakan jumlah output yang dihasilkan dalam periode tertentu dan berkaitan erat dengan tingkat permintaan pasar. Data penjualan

merupakan representasi dari permintaan konsumen yang dapat digunakan sebagai dasar dalam menentukan jumlah produksi yang akan dilakukan. Ketidaksesuaian antara data penjualan dan keputusan produksi berpotensi menimbulkan inefisiensi operasional, seperti penumpukan persediaan maupun kekurangan stok (Rahadiani et al., 2023).

Dalam konteks produksi tempe, data penjualan sering mengalami fluktuasi yang dipengaruhi oleh faktor musiman, kebiasaan konsumen, dan kondisi pasar. Oleh karena itu, pemanfaatan data penjualan historis dapat membantu produsen memperkirakan kebutuhan produksi secara lebih akurat. Pendekatan ini sejalan dengan konsep *data-driven Decision making*, yaitu pengambilan keputusan yang didasarkan pada analisis data yang objektif (Sugiarto et al., 2024). Dengan demikian, data penjualan memiliki peran penting dalam pembentukan model prediksi volume produksi karena mampu memberikan informasi mengenai pola permintaan yang dapat digunakan sebagai dasar dalam memperkirakan jumlah produksi pada periode berikutnya. Oleh sebab itu, data penjualan digunakan sebagai salah satu variabel input dalam penelitian ini.

2.3 Bahan Baku Produksi

Bahan baku adalah komponen utama dalam proses produksi yang digunakan sebagai input dasar untuk menghasilkan produk akhir. Dalam manajemen produksi, bahan baku didefinisikan sebagai material yang diolah melalui serangkaian proses untuk menghasilkan produk yang siap dikonsumsi. Ketersediaan bahan baku secara langsung mempengaruhi kelancaran proses produksi dan kapasitas output yang dapat dihasilkan. Pengelolaan bahan baku

yang tidak optimal dapat menyebabkan gangguan produksi, peningkatan biaya, dan penurunan mutu produk. Pentingnya perencanaan kebutuhan bahan baku yang akurat untuk menjaga stabilitas sistem produksi, terutama pada industri kecil dan menengah (Nasik & Pusakaningwati, 2025). Dalam konteks produksi tempe, pencatatan ini mencakup penggunaan kedelai sebagai bahan utama, ragi untuk fermentasi, serta bahan pendukung seperti air dan rempah yang mendukung kualitas akhir produk. Pengelolaan data bahan baku memberi pemahaman mendalam tentang kapasitas produksi yang dapat dicapai, memungkinkan UMKM untuk mempersiapkan kebutuhan bahan secara tepat guna mendukung kelancaran produksi harian. Tanpa data ini, risiko kekurangan bahan atau pemborosan dapat meningkat, terutama dalam skala kecil di mana fluktuasi pasokan sering terjadi.

Pada produksi tempe bahan baku seperti kedelai sebagai utama yang menjadi sumber protein, dan kualitasnya sangat mempengaruhi hasil akhir produk tempe. Selain itu, ragi tempe, yang berisi mikroorganisme *Rhizopus sp.*, berperan penting dalam menciptakan tekstur, aroma, dan cita rasa tempe. Kualitas dan dosis ragi akan mempengaruhi tingkat fermentasi dan konsistensi produk tempe (Kusuma & Widajanti, 2025). Selain kedelai dan ragi, bahan pendukung seperti air bersih dan bahan pengemas juga esensial untuk menjamin proses produksi yang bersih dan aman.

2.3.1 Ketersediaan Bahan Baku dan Hubungannya dengan Produksi

Ketersediaan bahan baku adalah salah satu faktor kunci dalam proses produksi, terutama untuk industri pangan yang berskala kecil. Tanpa pasokan bahan baku yang memadai, proses produksi tidak dapat dilakukan secara optimal,

meskipun terdapat permintaan pasar yang tinggi. Perencanaan kebutuhan bahan baku menjadi krusial dalam menentukan volume produksi yang dapat direalisasikan. Dengan memanfaatkan data historis penggunaan bahan baku, produsen dapat merencanakan volume produksi yang realistis sesuai dengan kondisi aktual, sesuai dengan pendekatan perencanaan berbasis data yang mengintegrasikan data bahan baku dan penjualan (Saputra & Apsari, 2024).

Dalam konteks produksi tempe, bahan baku seperti kedelai memiliki pengaruh yang sangat dominan terhadap volume produksi. Oleh karena itu, data bahan baku menjadi variabel penting dalam penelitian ini karena mencerminkan batas maksimum kapasitas produksi yang dapat dicapai.

2.4 Prediksi dalam Perencanaan Produksi

Prediksi adalah proses memperkirakan kejadian atau nilai di masa depan berdasarkan pola, hubungan, dan kecenderungan yang diobservasi dari data historis. Dalam konteks ilmiah, prediksi dilakukan dengan memanfaatkan metode analitis dan statistika untuk menghasilkan estimasi yang seakurat mungkin. Aktivitas prediksi merupakan komponen penting dalam pengambilan keputusan operasional karena menyediakan dasar kuantitatif untuk perencanaan sumber daya di masa mendatang. Berbeda dengan perkiraan subjektif, prediksi berbasis pada data dan model teruji, yang dapat mengurangi ketidakpastian dan risiko akibat perubahan kondisi (Moniz, 2022). Keputusan berbasis data melalui proses prediksi lebih konsisten dan akurat ketimbang model berbasis intuisi, terutama dalam sistem produksi yang dinamis (Taufik et al., 2021).

2.4.1 Peran Prediksi dalam Perencanaan Produksi

Perencanaan produksi melibatkan penentuan jumlah, waktu, dan cara produksi untuk memenuhi kebutuhan pasar secara efisien. Dalam konteks ini, prediksi permintaan memegang peranan sentral, membantu produsen dalam menyeimbangkan kapasitas produksi, ketersediaan bahan baku, dan tingkat persediaan yang optimal. Efektivitas perencanaan produksi sangat bergantung pada akurasi prediksi; kesalahan dalam prediksi dapat menyebabkan ketidaksesuaian pasokan dan permintaan, khususnya dalam industri pangan yang sering menghadapi batasan umur simpan produk (Raharusun, 2025). Kesalahan dalam prediksi permintaan bisa berakibat pada pemborosan jika prediksi terlalu tinggi, atau kekurangan stok dan kehilangan peluang penjualan jika prediksi terlalu rendah. Penerapan metode prediksi yang sederhana dan sistematis dapat meningkatkan kinerja perencanaan produksi (Pratama et al., 2025).

2.4.2 Prediksi Volume Produksi Berbasis Data

Pendekatan prediksi volume produksi berbasis data menggunakan data historis sebagai dasar untuk menentukan jumlah produksi di masa depan. Data yang digunakan dapat mencakup data penjualan, data bahan baku, serta variabel pendukung lainnya yang relevan. Dengan integrasi data ini, produsen dapat membangun model prediksi yang lebih objektif dan responsif terhadap perubahan kondisi pasar. Konsep ini mencerminkan ide pengambilan keputusan berbasis data, di mana keputusan operasional ditetapkan berdasarkan analisis data, bukan hanya berdasarkan pengalaman atau asumsi (Raharusun & Hasibuan, 2024). Dalam produksi tempe, pendekatan berbasis data ini sangat relevan untuk

mengatasi fluktuasi permintaan dan keterbatasan pasokan bahan baku. Model prediksi yang baik dapat membantu produsen menyesuaikan volume produksi sesuai dengan kondisi pasar, meningkatkan efisiensi penggunaan sumber daya, serta mengurangi risiko pemborosan (Febriyan, 2025). Dengan demikian, prediksi volume produksi berbasis data menjadi landasan penting dalam penelitian ini, terutama dalam penerapan algoritma *machine learning* seperti *Decision Tree* untuk pembangunan model prediksi yang akurat dan mudah diinterpretasikan.

2.5 Data Mining

Data Mining berfungsi menegaskan posisi algoritma *Decision Tree* sebagai salah satu teknik analitis yang bekerja di atas data historis penjualan dan bahan baku Tempe Asli HB. Istilah *Data Mining* merujuk pada penggunaan metode komputasional untuk mengekstraksi pola, hubungan, dan struktur tersembunyi dari himpunan data berukuran besar sehingga menghasilkan pengetahuan yang bermanfaat bagi pengambilan keputusan. Palacios et al. (2021) menjelaskan bahwa *Data Mining* menjadi inti proses penemuan pengetahuan ketika berbagai algoritma *Machine learning* digunakan guna mengidentifikasi faktor-faktor yang memengaruhi keberlanjutan studi mahasiswa berbasis data historis yang besar dan heterogen. Cara pandang ini sejalan dengan tujuan penelitian Tempe Asli HB, yaitu menggali pola antara penjualan, pemakaian bahan baku, dan volume produksi sehingga keputusan produksi memperoleh landasan empiris.

Istilah *Data Mining* sering diletakkan sebagai salah satu tahapan pada kerangka *Knowledge Discovery In Databases* (KDD). Tahap tersebut berfokus pada penerapan algoritma untuk menemukan pola dari data yang telah melalui

seleksi, pembersihan, dan transformasi. Berbeda dengan analisis deskriptif yang hanya menyajikan ringkasan angka, *Data Mining* menekankan pencarian struktur yang memiliki kemampuan prediktif maupun kemampuan untuk menjelaskan perilaku sistem. Menempatkan aktivitas pemodelan menggunakan algoritma seperti *Decision Tree* sebagai jantung proses penemuan pola yang kemudian diinterpretasikan bagi kepentingan kebijakan. Penelitian Tempe Asli HB memanfaatkan kerangka yang serupa ketika data penjualan dan bahan baku diproses melalui algoritma *Decision Tree* guna memperkirakan volume produksi tempe pada periode berikutnya.

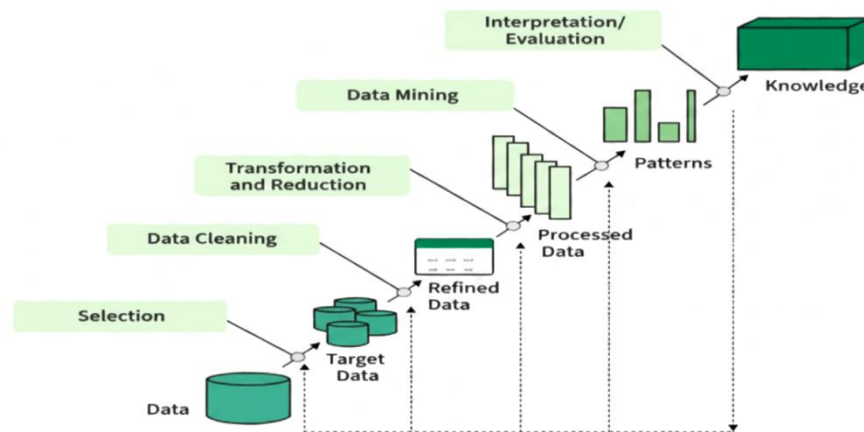
Karakter utama *Data Mining* terletak pada kemampuannya mendukung berbagai tugas analitis, antara lain klasifikasi, regresi, klasterisasi, asosiasi, maupun deteksi anomali. Pada ranah keuangan, Ataman dan Kahraman (2022) membandingkan kinerja *Decision Tree* dan aturan asosiasi untuk menginterpretasikan ekspektasi pelaku pasar modal terhadap kinerja indeks saham BIST100 dan BIST30. Penelitian tersebut memperlihatkan bahwa *Data Mining* dapat digunakan untuk menghubungkan variabel-variabel masukan dengan harapan investor, serta mengekstraksi aturan yang menjelaskan pola respon pasar. Pola pemikiran ini relevan bagi Tempe Asli HB karena data penjualan tempe dan pemakaian bahan baku juga menyimpan pola hubungan yang dapat diidentifikasi oleh algoritma *Decision Tree* sehingga menghasilkan model prediksi volume produksi.

2.6 *Knowledge Discovery In Databases*

Knowledge Discovery In Databases (KDD) adalah metodologi penting yang berfungsi untuk mengolah dan menganalisis data yang tersimpan dalam database. Proses ini memungkinkan peneliti dan praktisi untuk mengekstrak informasi baru yang lebih berarti dan mudah dipahami dari data yang ada. KDD diakui secara luas sebagai kerangka kerja sistematis yang mencakup beberapa tahap, termasuk pemilihan data, pembersihan, transformasi, penambahan data (*Data Mining*), serta interpretasi dan evaluasi hasil yang diperoleh (Putri et al., 2024). Dengan demikian, KDD bukan hanya sekadar alat, tetapi sebuah pendekatan ilmiah yang mengedepankan presisi dan validitas dalam pengambilan keputusan berbasis data. Peneliti berupaya memanfaatkan data historis penjualan dan bahan baku Tempe Asli HB memerlukan fondasi teoritis yang menjelaskan bagaimana data mentah diolah menjadi pengetahuan yang siap digunakan untuk pengambilan keputusan produksi. Pengertian KDD dan tahapan-tahapannya telah dikemukakan oleh banyak peneliti dalam bidang ini. KDD meliputi serangkaian proses yang terintegrasi dan merupakan ekstraksi non-trivial dari informasi yang berguna dari data (Martono & Padel, 2022).

Konsep KDD umumnya dipahami sebagai proses berjenjang yang bertujuan menemukan pola yang valid, baru, bermanfaat, dan mudah dipahami pada sekumpulan data berukuran besar. Tahapan tersebut meliputi Seleksi atau Pemilihan Data yang relevan, pemrosesan dan Pembersihan Data untuk mengatasi data hilang dan nilai pencilaan, transformasi agar data siap dianalisis, penerapan

metode *Data Mining* untuk membangun model, serta evaluasi dan interpretasi hasil guna menghasilkan pengetahuan yang siap dipakai pengambil keputusan.



Gambar 2.1 Tahapan Proses KDD

Berikut adalah penjelasan mendetail mengenai tahapan-tahapan dalam *Knowledge Discovery In Database* (KDD):

1. Seleksi atau Pemilihan Data

Pada fase ini, data yang relevan untuk analisis diidentifikasi dan dipilih. Pemilihan data yang tepat sangatlah penting karena kualitas dan relevansi data mempengaruhi hasil akhir dari proses KDD.

2. Pemrosesan dan Pembersihan Data

Pemrosesan dan pembersihan data adalah dua tahapan krusial dalam siklus *Knowledge Discovery In Database* (KDD). Proses ini bertujuan untuk memastikan bahwa data yang digunakan dalam analisis atau pemodelan tidak hanya lengkap dan konsisten, tetapi juga sesuai dan siap digunakan dalam model analisis. Tanpa pemrosesan dan pembersihan data yang tepat, hasil analisis atau model yang dibangun mungkin menyesatkan atau tidak valid.

3. Transformasi

Setelah data dibersihkan, Langkah selanjutnya adalah mentransformasikan data. Transformasi data bertujuan untuk mengubah data yang telah diseleksi dan dibersihkan kedalam format atau bentuk yang lebih sesuai untuk dianalisis lebih lanjut. Pada tahap ini, menggunakan Teknik tertentu, seperti *Data Mining*.

4. *Data Mining*

Data Mining adalah langkah utama KDD. Pada fase ini diterapkan teknik analisis algoritma untuk menemukan pola atau informasi tersembunyi dalam data. *Data Mining* melibatkan penggunaan berbagai teknik statistik dan pembelajaran mesin untuk menganalisis data dan menemukan hubungan atau pola yang sebelumnya tidak diketahui.

5. Evaluasi

Setelah pola atau model sudah ditemukan maka langkah selanjutnya adalah evaluasi. Pada fase ini model yang ditemukan diuji untuk mengevaluasi efektivitasnya dalam mendeskripsikan data hasil yang diinginkan. Evaluasi bertujuan untuk mengetahui apakah pola atau model yang ditemukan cukup valid dan layak digunakan dalam pengambilan keputusan.

2.7 *Machine learning*

Machine learning (ML) adalah cabang dari kecerdasan buatan yang mempelajari metode dan algoritma agar sistem dapat belajar dari data dan memperbaiki kinerjanya seiring bertambahnya pengalaman, tanpa pemrograman eksplisit untuk setiap kondisi (Sarker, 2021). ML memungkinkan komputer

mengenali pola, hubungan, dan kecenderungan dalam data historis untuk kemudian melakukan prediksi atau pengambilan keputusan otomatis. Karena volume dan kompleksitas data operasional meningkat, pendekatan ML dinilai lebih adaptif dibandingkan metode analisis konvensional sehingga banyak diterapkan pada bidang bisnis, kesehatan, pertanian, dan manufaktur termasuk perencanaan produksi.

Secara umum, *machine learning* diklasifikasikan menjadi tiga utama, yaitu :

1. *Supervised*

Supervised learning menggunakan data berlabel untuk mempelajari peta dari input ke output sehingga dapat melakukan prediksi (klasifikasi atau regresi) pada data baru.

2. *UnSupervised*

UnSupervised learning bekerja tanpa label, bertujuan menemukan struktur tersembunyi seperti kluster atau asosiasi dalam data.

3. *Semi-Supervised*

Semi-Supervised learning menggabungkan sebagian data berlabel dan sejumlah besar data tidak berlabel untuk meningkatkan performa model ketika pelabelan terbatas.

Pemilihan metode pembelajaran dalam *Machine learning* perlu disesuaikan dengan karakteristik data, termasuk ketersediaan label, serta tujuan analisis yang ingin dicapai. Secara umum, *Supervised learning* digunakan pada data berlabel, sedangkan *unSupervised learning* diterapkan pada data tanpa label (Sarker, 2021).

2.7.1 *Supervised learning* dalam Prediksi Produksi

Dalam penelitian ini, *Supervised learning* relevan untuk permasalahan prediksi volume produksi karena cocok untuk tugas regresi yang dimana dipergunakan untuk memprediksi nilai numerik seperti volume produksi aktual, data penjualan, penggunaan bahan baku yang dapat digunakan untuk melatih model prediktif. Dengan *Supervised learning*, algoritma mempelajari hubungan antara variabel input dan target output, sehingga dapat menghasilkan estimasi kuantitatif yang dapat dievaluasi menggunakan metrik akurasi atau error dan divalidasi secara ilmiah. Berbagai studi aplikasi industri menunjukkan dominasi metode *Supervised* untuk tugas-tugas prediksi dan klasifikasi dalam konteks produksi dan pemeliharaan peralatan.

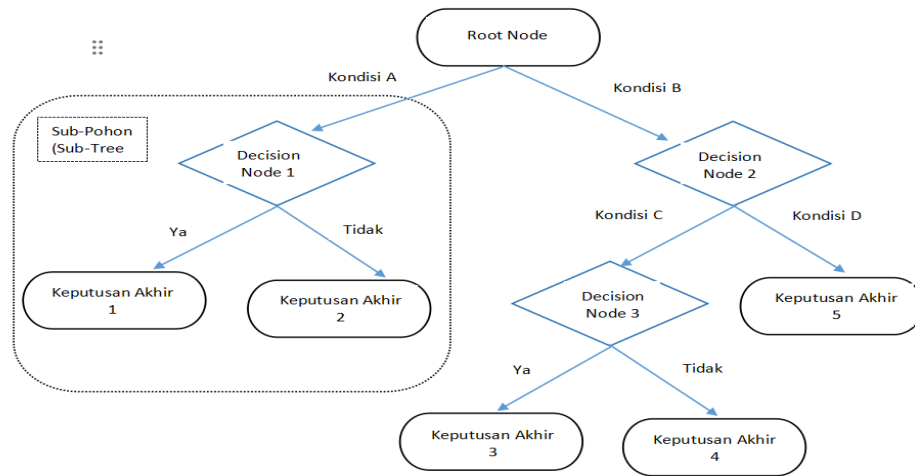
2.7.2 Peran *Machine learning* dalam Perencanaan Produksi

ML mendukung perencanaan produksi berbasis data dengan memanfaatkan data historis penjualan, bahan baku, dan variabel operasional untuk memprediksi permintaan dan kebutuhan produksi sehingga mengurangi ketergantungan pada intuisi (Atzeni et al., 2023). Pada usaha kecil dan menengah (UKM/UMKM), penelitian dan *dataset* terapan menunjukkan bahwa ML dapat memberikan wawasan berguna meskipun dengan data terbatas, bila diiringi teknik pra-proses dan pemilihan fitur yang tepat. Selain itu, model interpretatif seperti *Decision Tree* menawarkan keuntungan praktis untuk implementasi pada skala usaha kecil karena keluaran model mudah dipahami dan dapat langsung dipakai untuk aturan keputusan operasional.

machine learning dalam prediksi produksi telah merevolusi industri pangan, termasuk produksi tempe, dengan kemampuan menganalisis data historis untuk mengantisipasi output. Algoritma seperti *Decision Tree* membangun model yang memprediksi volume berdasarkan variabel seperti jumlah penjualan dan ketersediaan kedelai. Ini membantu UMKM menghindari overproduction atau underproduction, meningkatkan efisiensi.

2.8 Algoritma *Decision Tree* (*Regression Tree*)

Algoritma *Decision Tree Regression* merupakan metode *Supervised learning* yang merepresentasikan proses pengambilan keputusan dalam struktur pohon yang terdiri dari simpul (*node*), cabang (*branch*), dan daun (*leaf*) sehingga dapat digunakan untuk tugas klasifikasi maupun regresi (Charbuty & Abdulazeez, 2021). Karena pohon keputusan memetakan pengujian atribut pada simpul dan hasilnya pada cabang hingga mencapai daun yang memuat keputusan atau nilai output. Pada penelitian ini, *Decision Tree* digunakan untuk tugas regresi, yaitu memprediksi nilai numerik volume produksi tempe berdasarkan variabel input berupa data penjualan dan ketersediaan bahan baku. Varian yang umum digunakan untuk regresi adalah CART (*Classification and Regression Tree s*), yang membangun pohon berdasarkan pengurangan kesalahan prediksi pada setiap pemisahan data.



Gambar 2.2 Decision Tree

Struktur ini berbentuk hierarki seperti pohon terbalik, di mana setiap bagian memiliki peran tertentu dalam proses pengambilan keputusan.

1. *root node* (simpul akar). Ini merupakan titik awal dari seluruh proses. Pada tahap ini, data pertama kali diuji berdasarkan suatu kriteria tertentu untuk menentukan percabangan berikutnya. *root node* hanya ada satu dalam satu pohon keputusan.
2. Cabang (branch) adalah garis atau panah penghubung antar simpul (*node*). Secara lebih jelas: Cabang menghubungkan *root node* ke *Decision node*, menghubungkan *Decision node* ke *Decision node* lainnya, dan menghubungkan *Decision node* ke *leaf node*. Setiap cabang merepresentasikan hasil dari suatu pengujian atau kondisi tertentu. Misalnya, jika sebuah *Decision node* menguji suatu atribut, maka cabang yang keluar dari *node* tersebut menunjukkan kemungkinan hasilnya

3. *Decision node* (simpul keputusan) yang berada di tengah struktur. Simpul ini berfungsi untuk melakukan pengujian lanjutan terhadap atribut atau variabel tertentu. Setiap *Decision node* menghasilkan cabang-cabang yang merepresentasikan kemungkinan hasil dari pengujian tersebut. Dari sini, proses akan terus berlanjut hingga mencapai hasil akhir.
4. *leaf node* (simpul daun) merupakan bagian akhir. Simpul ini tidak memiliki percabangan lagi dan berisi hasil akhir dari proses pengambilan keputusan, seperti klasifikasi atau nilai prediksi.
5. Pada sisi kiri gambar, terdapat area yang diberi penanda *sub-Tree* (subpohon). *Sub-Tree* merupakan bagian dari pohon utama yang memiliki struktur serupa, yaitu terdiri dari *Decision node* dan *leaf node*. *Sub-Tree* ini menunjukkan bahwa setiap cabang dalam pohon dapat dianggap sebagai pohon kecil yang berdiri sendiri namun tetap menjadi bagian dari keseluruhan struktur.

Secara keseluruhan, gambar tersebut menjelaskan bagaimana sebuah pohon keputusan bekerja secara bertahap: dimulai dari akar, kemudian bercabang melalui beberapa simpul keputusan, hingga akhirnya mencapai simpul daun sebagai hasil akhir. Struktur ini membantu mempermudah proses analisis karena setiap langkah keputusan dapat ditelusuri dengan jelas dan sistematis.

2.8.1 Cara Kerja Algoritma *Decision Tree*

Algoritma *Decision Tree* bekerja dengan mekanisme pembentukan pohon secara rekursif dengan memilih atribut terbaik sebagai akar (root) berdasarkan atribut dan titik pemisah (*split point*) yang meminimalkan error pada *node*. Pada regresi, kriteria pemisahan umumnya menggunakan Mean Squared Error (MSE) atau variance reduction. *dataset* dibagi menjadi subset yang lebih homogen terhadap nilai target numerik hingga mencapai kondisi berhenti, seperti kedalaman maksimum atau jumlah minimum data pada *node*, misal semua contoh pada simpul memiliki kelas yang sama atau tidak ada atribut tersisa (Lotfi et al., 2021). Nilai prediksi pada *node* daun ditentukan sebagai rata-rata nilai target pada *node* tersebut.

Setelah fase pertumbuhan pohon (*growth*), sering dilakukan tahap pemangkasan (*pruning*) untuk mengurangi kompleksitas model dan mencegah *overfitting*; teknik *pruning* dapat berupa *pre-pruning* (menghentikan pertumbuhan lebih awal) atau *post-pruning* (memotong cabang setelah pohon terbentuk) yang terbukti meningkatkan generalisasi model pada data tidak terlihat (Lotfi et al., 2021).

Dalam konteks prediksi produksi, atribut seperti data penjualan dan ketersediaan bahan baku dapat diperlakukan sebagai fitur input; setiap cabang mewakili kondisi kombinasi atribut tersebut dan daun memberi nilai prediksi volume produksi atau aturan keputusan operasional yang siap diterapkan (Qolipour et al., 2021).

2.8.2 Kelebihan dan Kekurangan Algoritma *Decision Tree*

Kelebihan:

1. Interpretabilitas tinggi: *Decision Tree* menghasilkan aturan berbentuk if–then yang mudah dimengerti oleh pengguna non-teknis sehingga cocok untuk UKM yang memerlukan model yang dapat dijelaskan (Audemard et al., 2022).
2. Fleksibilitas data: Dapat menangani variabel numerik dan kategorikal tanpa transformasi kompleks, menjadikan *Decision Tree* praktis untuk data operasional yang heterogen seperti penjualan dan bahan baku (Charbuty & Abdulazeez, 2021).
3. Efisiensi komputasi relatif: Untuk *dataset* berukuran sedang, pembelajaran pohon umumnya lebih ringan dibandingkan beberapa metode *ensemble* atau *deep learning* sehingga implementasi pada lingkungan dengan sumber daya terbatas menjadi layak (Lotfi et al., 2021).

Kekurangan:

1. Sensitivitas terhadap perubahan data: *Decision Tree* dapat berubah drastis jika terdapat variasi kecil pada data, yang menunjukkan stabilitas model yang relatif rendah tanpa teknik tambahan seperti *ensemble* (Blanc et al., 2022).
2. Risiko overfitting: Pohon yang terlalu dalam dapat menyesuaikan diri pada noise data sehingga generalisasi buruk; oleh karena itu diperlukan pruning dan validasi silang untuk menjaga performa pada data baru (Lotfi et al., 2021).

3. Batasan optimalitas: Pembangunan pohon optimal secara komputasi bisa menjadi *intractable* (NP-hard) sehingga algoritma heuristik digunakan yang tidak selalu menjamin pohon paling ringkas secara teoritis (Blanc et al., 2022).

2.8.3 Penerapan Algoritma *Decision Tree* dalam Prediksi Produksi

Decision Tree telah diterapkan secara luas dalam tugas-tugas prediksi operasional dan manajemen risiko sehingga relevan untuk prediksi volume produksi tempe. Studi dan kasus terapan menunjukkan *Decision Tree* efektif untuk menghasilkan model yang mudah diinterpretasikan dan dapat dipadukan dengan proses pruning dan validasi untuk mengurangi overfitting (Lotfi et al., 2021). Dalam praktiknya, pipeline penerapan *Decision Tree* untuk prediksi produksi mencakup: pemilihan dan pra-proses fitur (mis. data penjualan historis, stok bahan baku, variabel musiman), pembagian data latih-uji, pelatihan pohon dengan kriteria pembelahan yang sesuai, pruning untuk mengendalikan kompleksitas, dan evaluasi menggunakan metrik kuantitatif serta pemeriksaan aturan yang dihasilkan untuk kebergunaan operasional (Putri et al., 2024).

Untuk skala usaha kecil seperti produsen tempe, keuntungan utama *Decision Tree* adalah keluaran aturan keputusan yang langsung dapat diterapkan (mis. “jika penjualan hari ini $> X$ dan stok kedelai $< Y$ maka produksi = Z ”), yang membantu membuat keputusan produksi harian menjadi lebih terukur dan transparan tanpa memerlukan pemahaman teknis yang mendalam (Audemard et al., 2022), (Qolipour et al., 2021).

2.8.4 Evaluasi Model *Decision Tree Regression*

Evaluasi model merupakan tahapan penting untuk menilai kinerja *Decision Tree Regression* dalam melakukan prediksi. Pada model regresi, evaluasi dilakukan dengan membandingkan nilai hasil prediksi dengan nilai aktual. Metrik evaluasi yang umum digunakan meliputi *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), serta koefisien determinasi (R^2).

Masing-masing metrik memiliki fungsi yang berbeda dalam menggambarkan tingkat kesalahan dan kemampuan model dalam menjelaskan variasi data. Penggunaan metrik evaluasi tersebut bertujuan untuk memastikan bahwa model yang dibangun memiliki tingkat akurasi yang baik dan dapat digunakan sebagai dasar pengambilan keputusan.

2.9 Tools Pendukung Penelitian

Tools pendukung penelitian merupakan perangkat lunak yang digunakan untuk membantu proses pengolahan data, analisis, serta pengujian model dalam penelitian ini. Penggunaan tools yang tepat bertujuan untuk meningkatkan ketelitian, efisiensi, dan keakuratan hasil penelitian. Adapun tools pendukung yang digunakan dalam penelitian ini adalah Microsoft Excel dan RapidMiner.

1. Microsoft Excel

Microsoft Excel merupakan perangkat lunak pengolah data berbasis spreadsheet yang banyak digunakan dalam kegiatan pengolahan dan analisis data. Dalam penelitian ini, Microsoft Excel berperan sebagai alat bantu pada tahap awal

pengolahan data sebelum dilakukan analisis lebih lanjut menggunakan algoritma *Data Mining*.

Microsoft Excel digunakan untuk melakukan input data, pembersihan data (data cleaning), serta pengelompokan data sesuai dengan kebutuhan penelitian. Proses pembersihan data meliputi pengecekan data ganda, penanganan data kosong, dan penyesuaian format data agar konsisten. Keunggulan penggunaan Microsoft Excel dalam penelitian ini adalah kemudahan dalam pengoperasian serta fleksibilitas dalam mengelola data numerik. Dengan fitur-fitur seperti fungsi statistik dasar, filter, dan pengurutan data, Excel membantu peneliti dalam memahami karakteristik data sebelum diterapkan ke dalam model *Decision Tree*.



Gambar 2.3 Aplikasi Microsoft Excel
Sumber : excelhelp.com

2. RapidMiner

RapidMiner merupakan perangkat lunak yang digunakan untuk analisis *Data Mining* dan *Machine learning* secara visual tanpa memerlukan penulisan kode. Dalam penelitian ini, RapidMiner dimanfaatkan untuk membangun model *Decision Tree Regression*, melakukan pembagian data latih dan uji, pruning, serta menghitung nilai evaluasi seperti MAE, MSE, RMSE, dan R^2 . Visualisasi pohon digunakan untuk mempermudah interpretasi aturan keputusan numerik.

Selain itu, RapidMiner mendukung proses *Knowledge Discovery In Databases* (KDD), mulai dari input data, pemilihan atribut, pemodelan, hingga evaluasi hasil. Data yang telah disiapkan menggunakan Microsoft Excel diimpor ke dalam RapidMiner untuk dianalisis menggunakan algoritma *Decision Tree*. Keunggulan RapidMiner terletak pada kemampuannya mengelola data numerik dan kategorikal secara efisien serta menyediakan visualisasi model yang mudah dipahami, sehingga membantu proses prediksi volume produksi tempe secara sistematis dan terukur.



Gambar 2.4 Aplikasi RapidMiner
Sumber : www.thrive.co.id

2.10 Penelitian terdahulu

Berikut 5 penelitian terdahulu yang digunakan sebagai dasar untuk memahami pendekatan, metode, dan hasil penelitian yang relevan dengan topik prediksi dan Algoritma serta hasil penelitiannya. Ringkasan penelitian terdahulu yang berkaitan dengan penelitian ini disajikan pada Tabel berikut.

Tabel 2.1 Penelitian Terdahulu

No	Peneliti	Judul Penelitian	Tahun	Data dan Algoritma	Hasil Penelitian
1	Dwi Septiani et al.	Implementasi <i>Decision Tree</i> untuk Prediksi	2025	Data historis produksi	Model yang dioptimasi menunjukkan

		Kebutuhan Bahan Kain pada Usaha Konveksi		celana, <i>Decision Tree</i> Regressor	performa lebih baik, dengan kesalahan absolut rata-rata sebesar 0,7851 yard dan mampu menjelaskan 98,62% variabilitas data. Hasil penelitian menunjukkan bahwa model yang diusulkan dapat meningkatkan efisiensi manajemen stok bahan kain.
2	Suhartini et al.	Klasifikasi Penjualan Tempe dengan <i>Data Mining</i> Menggunakan Algoritma C4.5	2023	Data penjualan tempe, Algoritma C4.5	Klasifikasi penjualan tempe pada usaha Bapak Suswanto terbukti dengan mendapatkan hasil akurasi sebesar 69.72%, Classification precision sebesar 71.20% dan AUC sebesar 0,586.
3	Cecep Maulana Sidiq et al.	Algoritma <i>Decision Tree</i> C4.5 untuk Mengklasifikasikan Data Stroke	2024	Data stroke, Algoritma C4.5	Temuan fase evaluasi mengungkapkan tingkat akurasi yang mengesankan 93,64%, dengan nilai presisi 12,50% untuk prediksi stroke positif dan 95,26% untuk prediksi stroke negatif.

4	Abdul Halim Anshor et al.	Implementasi <i>Decision Tree</i> pada Sistem Prediksi Status Kualitas Produk Minuman A	2024	Data kualitas minuman, <i>Decision Tree</i>	Hasil penelitian ini akan menunjukkan presentasi nilai akurasi dari kualitas produk A sebesar 99.59% .
5	Gunawan	Penerapan <i>Data Mining</i> untuk Memprediksi Harga Udang Vaname di Pasar Lokal Menggunakan Algoritma <i>Decision Tree</i>	2025	Data harga udang vaname, <i>Decision Tree</i>	Akurasi 88.89%, faktor cuaca dan harga pakan paling berpengaruh
6	Penelitian ini	Penerapan Algoritma <i>Decision Tree</i> untuk Prediksi Volume Produksi Tempe Asli Hb berdasarkan Data Penjualan dan Bahan Baku	2026	Data penjualan dan bahan baku tempe; algoritma <i>Decision Tree</i>	Penelitian ini mengambil posisi yang berbeda dengan memfokuskan pada prediksi numerik volume produksi menggunakan algoritma <i>Decision Tree Regression</i> serta mengintegrasikan data penjualan dan bahan baku sebagai variabel input.

Penelitian terdahulu menunjukkan bahwa algoritma *Decision Tree* banyak digunakan dalam berbagai bidang, baik untuk klasifikasi maupun regresi. Namun, sebagian besar penelitian masih berfokus pada klasifikasi, seperti klasifikasi kualitas produk atau kategori penjualan.

Berbeda dengan penelitian tersebut, penelitian ini berfokus pada prediksi numerik volume produksi menggunakan pendekatan *Decision Tree Regression*

dengan mengintegrasikan data penjualan dan bahan baku. Hal ini menjadi nilai kebaruan karena menggabungkan dua variabel utama dalam konteks UMKM serta menghasilkan model yang bersifat prediktif dan aplikatif.

Dari sisi algoritma, *Decision Tree* dipilih bukan hanya karena akurasinya, tetapi karena kemampuannya menghasilkan aturan keputusan yang mudah dipahami oleh pelaku UMKM, sehingga hasil penelitian tidak berhenti pada evaluasi model, tetapi dapat langsung digunakan dalam perencanaan produksi.