

Analisis Penjualan pada Perusahaan Food and Beverage Menggunakan Algoritma Decision Tree dalam Meningkatkan Penjualan

Muhammad Ayudi^{1,*}, Bayu Rianto², Fitri Yunita³

^{1, 2, 3}Fakultas Teknik dan Ilmu Komputer, Sistem Informasi, Universitas Islam Indragiri, Tembilahan, Indonesia
Email: ^{1,*}masyudialfarisi@gmail.com, ²rianto.bayu91@gmail.com, ³fitriyun@gmail.com
^{*}Email Penulis Utama

Abstrak—Industri Food and Beverage (F&B) mengalami pertumbuhan yang pesat seiring dengan meningkatnya konsumsi masyarakat dan dinamika persaingan bisnis yang semakin kompleks. Kondisi tersebut menuntut perusahaan F&B untuk mampu memanfaatkan data penjualan secara optimal sebagai dasar pengambilan keputusan strategis. Data penjualan tidak hanya merepresentasikan kinerja bisnis, tetapi juga mencerminkan perilaku konsumen, tren permintaan, serta efektivitas strategi pemasaran yang diterapkan. Namun, dalam praktiknya, banyak perusahaan F&B masih belum memanfaatkan data penjualan secara analitis dan sistematis, sehingga strategi penjualan yang dihasilkan cenderung kurang responsif terhadap perubahan pasar. Penelitian ini bertujuan untuk menganalisis data penjualan perusahaan Food and Beverage (F&B) menggunakan algoritma Decision Tree guna meningkatkan strategi penjualan melalui identifikasi faktor-faktor yang memengaruhi tingkat penjualan. Metode penelitian yang digunakan adalah penelitian kuantitatif dengan pendekatan data mining, di mana data penjualan historis diolah melalui tahapan preprocessing, pemodelan, dan evaluasi menggunakan algoritma Decision Tree. Model yang dihasilkan dianalisis untuk mengetahui pola, hubungan antar variabel, serta aturan keputusan yang terbentuk. Hasil penelitian menunjukkan bahwa algoritma Decision Tree mampu mengklasifikasikan tingkat penjualan dengan baik serta mengidentifikasi variabel-variabel yang memiliki pengaruh signifikan terhadap penjualan, seperti kategori produk, harga, dan waktu penjualan. Temuan ini memberikan dasar analitis yang kuat bagi perusahaan dalam merumuskan strategi penjualan yang lebih tepat sasaran dan berbasis data. Secara akademik, penelitian ini berkontribusi pada pengembangan kajian data mining dalam konteks industri F&B, sementara secara praktis memberikan rekomendasi strategis yang dapat dimanfaatkan oleh manajemen perusahaan untuk meningkatkan kinerja penjualan.

Kata Kunci: Data Mining, Decision Tree, Penjualan, Food and Beverage (F&B), Strategi Penjualan, Analisis Data

Abstract—The Food and Beverage (F&B) industry is experiencing rapid growth along with increasing consumer consumption and increasingly complex business competition dynamics. This situation demands that F&B companies optimally utilize sales data as a basis for strategic decision-making. Sales data not only represents business performance but also reflects consumer behavior, demand trends, and the effectiveness of implemented marketing strategies. However, in practice, many F&B companies still do not utilize sales data analytically and systematically, resulting in sales strategies that tend to be less responsive to market changes. This study aims to analyze sales data from Food and Beverage (F&B) companies using the Decision Tree algorithm to improve sales strategies by identifying factors that influence sales levels. The research method used is quantitative research with a data mining approach, where historical sales data is processed through preprocessing, modeling, and evaluation using the Decision Tree algorithm. The resulting model is analyzed to identify patterns, relationships between variables, and the resulting decision rules. The results show that the Decision Tree algorithm is capable of effectively classifying sales levels and identifying variables that significantly influence sales, such as product category, price, and sales timing. These findings provide a strong analytical basis for companies to formulate more targeted, data-driven sales strategies. Academically, this research contributes to the development of data mining studies in the F&B industry, while practically, it provides strategic recommendations that company management can utilize to improve sales performance.

Keywords: Data Mining, Decision Tree, Sales, Food and Beverage (F&B), Sales Strategy, Data Analysis, Classification

1. PENDAHULUAN

Industri *Food and Beverage* (F&B) merupakan salah satu sektor bisnis yang mengalami pertumbuhan pesat seiring dengan perkembangan ekonomi, perubahan gaya hidup, dan meningkatnya konsumsi masyarakat di era modern. Sektor ini memiliki peran ganda yang strategis, tidak hanya sebagai penyedia kebutuhan dasar pangan tetapi juga sebagai salah satu pilar penting dalam perekonomian nasional. Melalui penciptaan lapangan kerja, peningkatan produktivitas, serta kontribusi terhadap Produk Domestik Bruto (PDB), industri F&B turut mendorong pertumbuhan ekonomi secara signifikan. Berdasarkan data terbaru dari Badan Pusat Statistik, nilai penjualan usaha penyediaan makanan dan minuman di Indonesia mencapai Rp998,37 triliun pada tahun 2023, mengalami peningkatan yang signifikan dibandingkan periode sebelumnya. Selain itu, jumlah unit usaha di sektor ini meningkat menjadi 4,85 juta, yang mampu menyerap sekitar 9,8 juta tenaga kerja [1]. Pertumbuhan ini menunjukkan bahwa industri F&B bukan hanya sektor yang berkembang karena tren konsumen, tetapi juga menjadi motor ekonomi yang berdampak luas terhadap kesejahteraan masyarakat.

Perkembangan industri makanan dan minuman (F&B) saat ini sangat dipengaruhi oleh perubahan lingkungan pasar yang dinamis serta tingkat persaingan yang semakin intens. Berbagai faktor seperti pergeseran pola konsumsi masyarakat, meningkatnya kesadaran konsumen terhadap aspek kualitas, keamanan, dan kesehatan

produk, serta kemajuan teknologi digital yang mempermudah akses informasi, telah mendorong perusahaan untuk bertindak lebih adaptif dan responsif. Kondisi ini menuntut pelaku usaha untuk dapat mampu merumuskan dan mengeksekusi strategi bisnis secara cepat, tepat, dan berbasis pada data yang lebih akurat dan efektif.

Dalam situasi tersebut, kemampuan perusahaan untuk mampu memahami tren pasar serta mampu menganalisis perilaku konsumen menjadi elemen krusial dalam menentukan keberhasilan bisnis. Perusahaan tidak lagi cukup hanya mengandalkan intuisi atau pengalaman semata, melainkan harus didukung oleh pemanfaatan data yang relevan dan terukur. Salah satu sumber informasi yang sangat penting dalam hal ini adalah data penjualan. Data penjualan tidak hanya berfungsi sebagai alat untuk mengukur kinerja bisnis secara keseluruhan, tetapi juga dapat memberikan wawasan yang lebih mendalam terkait preferensi pelanggan, pola pembelian, efektivitas program promosi, kinerja distribusi, hingga tingkat penerimaan pasar terhadap produk atau strategi tertentu yang diterapkan. Penjualan merupakan indikator sentral dalam evaluasi performa bisnis karena berhubungan langsung dengan pendapatan dan pertumbuhan perusahaan [2].

Oleh karena itu, pengelolaan dan analisis data penjualan secara terstruktur, sistematis, dan berkelanjutan menjadi kebutuhan yang tidak dapat diabaikan oleh perusahaan F&B. Dengan memanfaatkan analisis data yang tepat, perusahaan dapat menghasilkan *insight* yang bernilai untuk mendukung pengambilan keputusan strategis yang lebih objektif dan berbasis bukti (*data-driven decision making*). Hal ini pada akhirnya dapat membantu perusahaan dalam meminimalkan risiko kesalahan dalam menentukan strategi bisnis, sekaligus meningkatkan daya saing dan keberlanjutan usaha di tengah persaingan industri yang semakin ketat.

Namun, meskipun data penjualan memiliki potensi yang besar, banyak perusahaan F&B masih menghadapi berbagai tantangan dalam memanfaatkannya secara optimal. Salah satu permasalahan yang paling umum adalah ketidakmampuan memprediksi tren penjualan secara akurat. Permasalahan ini sering terjadi karena perusahaan masih mengandalkan pendekatan tradisional yang bersifat intuitif, tanpa memanfaatkan metode analisis yang mampu menangkap pola tersembunyi dalam data historis. Selain itu, kurangnya sistem informasi yang tersentralisasi menyebabkan data penjualan tersebar dan sulit diakses secara konsisten, sehingga analisis yang dilakukan seringkali tidak komprehensif.

Di samping itu, keterbatasan infrastruktur teknologi informasi juga menjadi hambatan signifikan. Banyak perusahaan belum memiliki sistem data yang terintegrasi secara menyeluruh, sehingga data penjualan tersebar di berbagai sumber dan sulit diakses secara konsisten. Kondisi ini menyebabkan proses analisis menjadi tidak menyeluruh dan berpotensi menghasilkan kesimpulan yang kurang akurat. Terakhir, keterbatasan penggunaan metode analitik canggih membuat banyak *insight* yang seharusnya dapat diperoleh tetap tersembunyi. Fenomena-fenomena ini menunjukkan bahwa perusahaan membutuhkan pendekatan yang lebih sistematis dan berbasis data untuk meningkatkan kualitas pengambilan keputusan, sehingga strategi penjualan dapat lebih responsif terhadap perubahan pasar dan perilaku konsumen.

Fenomena tersebut menunjukkan adanya kebutuhan mendesak bagi perusahaan untuk mengadopsi pendekatan yang lebih terstruktur, sistematis, dan berbasis data dalam proses pengambilan keputusan. Dengan pendekatan ini, perusahaan dapat merumuskan strategi penjualan yang lebih adaptif dan responsif terhadap perubahan kondisi pasar serta dinamika perilaku konsumen yang terus berkembang.

Dalam konteks tersebut, pendekatan data mining menjadi salah satu solusi yang relevan. Data mining merupakan teknik yang memungkinkan perusahaan untuk mengekstraksi informasi berharga dari dataset besar, mengenali pola, serta memprediksi tren di masa depan. Salah satu algoritma data mining yang banyak digunakan adalah *Decision Tree*. Algoritma ini memiliki kemampuan untuk mengklasifikasikan data dan memprediksi outcome berdasarkan berbagai atribut, serta menghasilkan model pohon keputusan yang mudah dipahami dan diinterpretasikan. Dalam industri F&B, atribut yang dianalisis bisa meliputi volume penjualan harian, mingguan atau bulanan, kategori produk, harga, promosi, lokasi outlet, hingga variabel eksternal seperti musim, hari libur, atau tren musiman tertentu. Keunggulan utama *Decision Tree* terletak pada kemampuannya menangkap hubungan non-linier antarvariabel dan menyoroti faktor-faktor yang paling berpengaruh terhadap tingkat penjualan.

Dengan demikian, penerapan algoritma ini tidak hanya membantu perusahaan dalam menghasilkan prediksi yang lebih akurat, tetapi juga memberikan pemahaman yang lebih komprehensif mengenai faktor-faktor kunci yang memengaruhi performa bisnis. Hal ini pada akhirnya dapat mendukung manajemen dalam merumuskan strategi yang lebih efektif, tepat sasaran, dan berbasis pada bukti empiris.

Berbagai penelitian terdahulu telah membuktikan efektivitas *Decision Tree* dalam konteks penjualan. Penelitian Smith dan Rodriguez membuktikan bahwa atribut promosi dalam model *Decision Tree* berperan signifikan dalam mengklasifikasikan pola pembelian pelanggan ritel [3]. Penelitian sebelumnya menunjukkan bahwa algoritma *Decision Tree* mampu mengidentifikasi kategori produk yang laris dan tidak laris dengan akurasi yang cukup tinggi dibandingkan metode lain, seperti *Random Forest* [4]. Selain itu, penelitian yang diterapkan pada perusahaan Mayora Group menunjukkan bahwa penggunaan *Decision Tree* dapat menghasilkan prediksi penjualan yang lebih akurat dan membantu perencanaan strategi pemasaran [5]. Penelitian lain membandingkan *Decision Tree* dengan metode *Naive Bayes* pada dataset penjualan minuman kopi, di mana *Decision Tree* menunjukkan kemampuan lebih baik dalam klasifikasi tingkat penjualan [6]. Penelitian lain membandingkan kinerja *Decision Tree* dengan *Random Forest* untuk klasifikasi data penjualan produk pada dataset ritel, di mana

Decision Tree menunjukkan kemampuan cukup tinggi dalam mengidentifikasi kategori produk yang laris dan tidak laris [7]. Meskipun demikian, sebagian besar penelitian sebelumnya masih berfokus pada dataset ritel umum atau kasus kecil tertentu, dan masih jarang yang secara komprehensif mengeksplorasi penggunaan Decision Tree untuk analisis data penjualan perusahaan F&B secara strategis.

Keterbatasan penelitian terdahulu dalam mengkaji pemanfaatan data penjualan secara komprehensif membuka peluang yang signifikan bagi penelitian ini untuk memberikan kontribusi yang lebih mendalam. Analisis yang dilakukan tidak hanya berfokus pada pengujian tingkat akurasi model Decision Tree, tetapi juga diarahkan untuk mengaitkan hasil analisis tersebut dengan implementasi strategi bisnis yang nyata dan aplikatif. Dalam konteks ini, berbagai variabel yang memengaruhi penjualan seperti faktor musiman, aktivitas promosi, lokasi outlet, serta dinamika perilaku konsumen perlu dikaji secara terintegrasi. Pendekatan yang menyeluruh ini diharapkan mampu menghasilkan rekomendasi strategi yang lebih tepat sasaran dan relevan dengan kondisi pasar yang aktual. Oleh karena itu, penelitian ini tidak hanya memberikan kontribusi pada pengembangan literatur akademik di bidang data mining dan analisis penjualan, tetapi juga menawarkan nilai praktis yang tinggi bagi pelaku industri F&B.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis data penjualan pada perusahaan F&B dengan memanfaatkan algoritma Decision Tree sebagai metode utama. Fokus penelitian meliputi beberapa aspek penting, yaitu mengidentifikasi faktor-faktor yang berpengaruh terhadap tingkat penjualan, mengevaluasi performa model dalam mengklasifikasikan kategori penjualan, serta merumuskan rekomendasi strategi yang didasarkan pada hasil analisis data. Dari sisi akademik, penelitian ini diharapkan dapat memperkaya kajian mengenai penerapan teknik data mining dalam sektor F&B, khususnya dalam konteks analisis penjualan. Sementara itu, dari sisi praktis, hasil penelitian ini diharapkan dapat menjadi panduan bagi manajemen perusahaan dalam mengoptimalkan pemanfaatan data sebagai dasar pengambilan keputusan strategis.

Dengan demikian, penelitian ini hadir sebagai upaya untuk menjembatani kesenjangan antara besarnya potensi data penjualan yang dimiliki perusahaan dan masih terbatasnya pemanfaatan data tersebut dalam praktik bisnis. Hasil yang diperoleh diharapkan tidak hanya memberikan insight yang kuat untuk keperluan prediksi penjualan, tetapi juga mampu mendukung proses pengambilan keputusan manajerial yang lebih berbasis data, adaptif terhadap perubahan pasar, serta berorientasi pada pencapaian target bisnis yang lebih optimal dan berkelanjutan.

2. METODE PENELITIAN

2.1 Tahapan Penelitian

Menurut Creswell, penelitian deskriptif kuantitatif bertujuan untuk menyajikan gambaran akurat mengenai fenomena yang diteliti berdasarkan data empiris [8]. Penelitian ini menggunakan pendekatan kuantitatif dengan memanfaatkan metode data mining melalui penerapan algoritma Decision Tree untuk menganalisis data penjualan pada perusahaan Food and Beverage (F&B) [9]. Pendekatan ini di ambil karena mampu mengolah data dalam jumlah besar secara sistematis serta mampu menghasilkan pola dan informasi yang dapat digunakan sebagai dasar dalam memperoleh suatu keputusan. Decision tree adalah proses yang dapat mengidentifikasi berbagai model kelas yang terpisah dan membedakan antar kelas. Decision tree sebuah metode machine learning yang didasarkan pada sebuah pohon keputusan yang dapat mengklasifikasi data dan Sebagian besar merupakan tipe data nominal [10]. Algoritma yang diterapkan dalam decision tree adalah algoritma C4.5 [11].

Algoritma C4.5 dipahami sebagai sebuah algoritma yang berfungsi untuk mengembangkan pohon keputusan. Konsep berpikir pohon keputusan dimulai dari atribut paling atas yang disebut sebagai akar (*root*) kemudian mengalir melalui cabang-cabang di bawahnya sehingga berakhir di atribut paling bawah yang disebut sebagai daun (*leaf*) [12]. Tahapan penelitian disusun secara sistematis untuk menghasilkan model analisis yang akurat dan sesuai dengan tujuan penelitian.

Tahap pertama adalah pengumpulan data, yaitu mengumpulkan data penjualan historis perusahaan F&B yang mencakup atribut seperti jenis produk, kategori produk, harga, jumlah penjualan, dan waktu transaksi [13]. Atribut-atribut ini dipilih karena dianggap mampu merepresentasikan kondisi penjualan secara menyeluruh dan menjadi fondasi utama sehingga, data ini menjadi dasar utama dalam proses analisis.

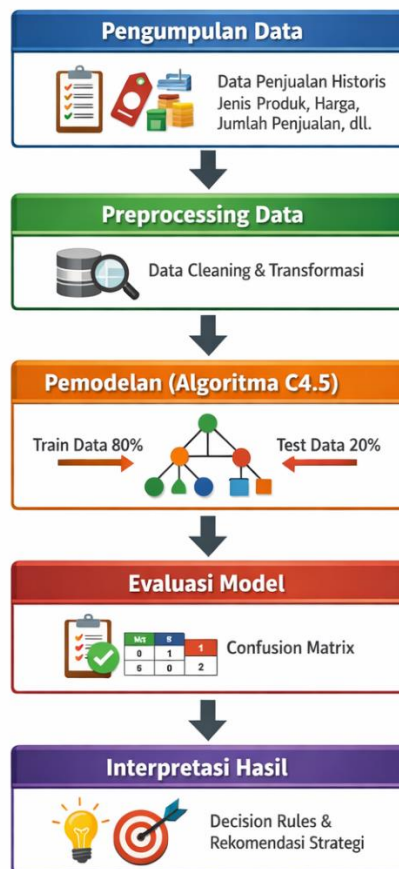
Tahap kedua adalah *preprocessing data*, yang bertujuan untuk meningkatkan kualitas dan konsistensi data sebelum digunakan dalam pemodelan. Tahapan ini meliputi beberapa proses penting, antara lain pembersihan data (*data cleaning*), penghapusan data duplikat, penanganan nilai kosong (*missing value*), serta transformasi data ke dalam format yang sesuai untuk proses pemodelan. Tahap ini sangat krusial karena kualitas data yang baik akan sangat memengaruhi performa model yang dihasilkan.

Tahap ketiga adalah pemodelan dengan menggunakan algoritma Decision Tree. Pada tahap ini, data yang telah melalui proses *preprocessing* digunakan untuk membangun model klasifikasi. Model tersebut bertujuan untuk mengelompokkan tingkat penjualan ke dalam kategori tertentu berdasarkan atribut-atribut yang tersedia. Algoritma Decision Tree dipilih karena kemampuannya dalam menghasilkan model yang mudah dipahami

melalui struktur pohon keputusan, serta kemampuannya dalam mengidentifikasi hubungan antarvariabel secara efektif.

Tahap keempat adalah evaluasi model untuk mengukur kinerja model yang dihasilkan. Evaluasi ini dilakukan dengan menggunakan beberapa metrik, seperti akurasi untuk mengetahui tingkat ketepatan prediksi model, serta confusion matrix untuk melihat performa klasifikasi secara lebih rinci, termasuk tingkat kesalahan dalam pengelompokan data. Melalui tahap ini, dapat diketahui sejauh mana model mampu merepresentasikan pola dalam data penjualan secara akurat.

Tahap terakhir adalah interpretasi hasil, yaitu menganalisis pola serta aturan keputusan (decision rules) yang dihasilkan model sebagai dasar dalam penyusunan strategi penjualan berbasis data. yaitu proses menganalisis output dari model Decision Tree. Pada tahap ini, peneliti mengkaji pola-pola yang terbentuk serta aturan keputusan (decision rules) yang dihasilkan oleh model. Hasil interpretasi ini kemudian digunakan sebagai dasar dalam merencanakan strategi penjualan yang lebih efektif dan berbasis data. Dengan memahami faktor-faktor yang memengaruhi tingkat penjualan, perusahaan dapat mengambil langkah strategis yang lebih tepat, seperti penyesuaian harga, optimalisasi promosi, maupun pengelolaan produk yang lebih efisien. Tahapan penelitian yang dilakukan dalam penelitian ini dapat dilihat pada Gambar 1. berikut:



Gambar 1. Alur Proses Data Mining Menggunakan Decision Tree

2.2 Variabel Penelitian

Dalam penelitian ini, variabel yang digunakan terbagi menjadi dua kelompok utama, yaitu variabel independen dan variabel dependen, yang masing-masing memiliki peran penting dalam proses analisis. Variabel independen adalah atribut-atribut dari data penjualan yang berfungsi sebagai faktor penentu atau predictor dalam model Decision Tree. Variabel-variabel ini merepresentasikan karakteristik transaksi penjualan yang dapat memengaruhi hasil penjualan, baik secara langsung maupun tidak langsung. Contoh variabel independen meliputi jenis produk, kategori produk, harga produk, jumlah penjualan, waktu transaksi, metode pembayaran, dan saluran penjualan. Dalam algoritma Decision Tree, variabel independen digunakan untuk membangun struktur pohon keputusan melalui proses pemilihan atribut terbaik pada setiap percabangan, sehingga model dapat mengidentifikasi faktor-faktor yang paling berpengaruh terhadap klasifikasi kinerja penjualan.

Sementara itu, variabel dependen merupakan output atau target yang ingin diprediksi oleh model. Dalam konteks penelitian ini, variabel dependen adalah kinerja penjualan, yang dikategorikan ke dalam tingkatan

tertentu, misalnya penjualan tinggi, sedang, atau rendah. Variabel dependen berfungsi sebagai hasil akhir dari proses klasifikasi dan menjadi indikator utama untuk menilai performa bisnis.

Dengan pemisahan kedua jenis variabel ini, model Decision Tree dapat bekerja secara efektif: variabel independen menyediakan informasi yang dibutuhkan untuk membentuk aturan keputusan, sedangkan variabel dependen menjadi acuan untuk mengevaluasi akurasi prediksi dan mengekstrak insight yang dapat digunakan dalam pengambilan keputusan strategis perusahaan.

2.2.1 Variabel Independen (X)

Variabel independen Variabel independen dalam penelitian ini adalah atribut-atribut dari data penjualan yang berperan sebagai faktor penentu atau predictor dalam model klasifikasi. Variabel-variabel ini digunakan untuk menjelaskan atau memengaruhi variabel dependen, yaitu kinerja penjualan, sehingga dapat membantu algoritma Decision Tree dalam membentuk pola dan struktur pohon keputusan. Setiap variabel independen merepresentasikan aspek tertentu dari transaksi penjualan yang secara langsung maupun tidak langsung berkontribusi terhadap hasil penjualan.

Dalam penelitian ini, variabel independen yang digunakan meliputi beberapa atribut penting, Secara konseptual, atribut penjualan merepresentasikan karakteristik transaksi yang memengaruhi kinerja penjualan [14]. Variabel ini meliputi:

- a. jenis produk
- b. kategori produk
- c. harga produk
- d. jumlah penjualan
- e. waktu transaksi
- f. metode pembayaran
- g. saluran penjualan

Variabel ini berperan sebagai predictor attributes dalam proses klasifikasi. Variabel-variabel tersebut merepresentasikan karakteristik transaksi penjualan yang secara langsung maupun tidak langsung memengaruhi kinerja penjualan. Setiap atribut memiliki peran dalam membentuk pola hubungan antar data yang akan dianalisis. Dalam algoritma Decision Tree, variabel independen berfungsi sebagai predictor attributes, yaitu variabel yang digunakan untuk membangun struktur pohon keputusan melalui proses pemilihan atribut terbaik pada setiap percabangan. Hasil dari proses ini akan menunjukkan atribut mana yang paling dominan dalam menentukan klasifikasi kinerja penjualan

2.2.2 Variabel Dependen (Y)

Variabel dependen dalam penelitian ini berfungsi sebagai hasil atau target klasifikasi yang ingin diprediksi oleh model. Dengan kata lain, variabel ini merupakan output dari proses analisis yang mencerminkan kinerja penjualan perusahaan F&B. Variabel dependen dipilih karena mencerminkan indikator utama keberhasilan bisnis dan menjadi fokus utama dalam pengambilan keputusan strategis. Secara teoritis, kinerja penjualan mencerminkan tingkat keberhasilan perusahaan dalam mencapai target penjualan dan memaksimalkan potensi pasar [15]. Dalam penelitian ini, kinerja penjualan direpresentasikan dalam bentuk kategori yang dapat diklasifikasikan, yaitu kinerja penjualan yang direpresentasikan dalam bentuk:

- a. tingkat penjualan
- b. status kinerja (tercapai / tidak tercapai)
- c. pencapaian target

Variabel ini menjadi class label dalam algoritma Decision Tree. Ketiga indikator tersebut digunakan untuk menggambarkan tingkat keberhasilan penjualan dalam periode tertentu berdasarkan data historis yang tersedia. Variabel ini kemudian diklasifikasikan ke dalam kategori tertentu sehingga dapat diolah dalam proses pemodelan. Dalam algoritma Decision Tree, variabel dependen berperan sebagai class label, yaitu variabel target yang akan diprediksi berdasarkan atribut-atribut pada variabel independen. Hasil klasifikasi ini selanjutnya digunakan untuk mengidentifikasi pola penjualan dan mendukung pengambilan keputusan strategis perusahaan.

Selain itu, penggunaan variabel dependen dalam bentuk kategorikal ini juga memudahkan proses interpretasi hasil model, karena output yang dihasilkan tidak lagi berupa nilai numerik yang kompleks, melainkan berupa kelas atau status yang lebih mudah dipahami. Hal ini memungkinkan pihak manajemen untuk dengan cepat mengevaluasi kondisi kinerja penjualan dalam periode tertentu serta menentukan langkah-langkah perbaikan atau strategi yang lebih tepat sasaran berdasarkan hasil klasifikasi yang diperoleh dari model Decision Tree.

Tabel 1. Definisi Operasional Variabel

No	Variabel	Jenis	Indikator	Skala
1	Atribut Data Penjualan	Independen	Jenis Produk	Nominal
			Kategori Produk	Nominal
			Harga Produk	Rasio
			Jumlah Penjualan	Rasio
			Waktu Transaksi	Ordinal
			Metode Pembayaran	Nominal
2	Kinerja Penjualan	Dependen	Saluran Penjualan	Nominal
			Tingkat Penjualan	Ordinal
			Status Kinerja	Ordinal
			Pencapaian Target	Rasio/Ordinal

2.3 Teknik Pengumpulan Data

Dalam penelitian ini, teknik pengumpulan data dilakukan melalui tiga pendekatan utama, yaitu studi dokumentasi, kuesioner, dan studi pustaka, masing-masing memiliki peran dan tujuan yang spesifik dalam mendukung analisis.

Studi dokumentasi digunakan untuk memperoleh data primer yang bersumber langsung dari perusahaan F&B. Menurut Sugiyono, data primer memiliki tingkat akurasi yang tinggi karena diperoleh langsung dari sumbernya dan relevan secara langsung dengan tujuan penelitian [16]. Data ini diperoleh dari sistem informasi perusahaan yang merekam seluruh transaksi penjualan secara digital. Informasi yang dikumpulkan mencakup berbagai atribut penting, antara lain jenis produk, kategori produk, harga jual, jumlah unit terjual, waktu transaksi, serta atribut tambahan yang berkaitan dengan proses penjualan. Data historis ini menjadi fondasi utama bagi analisis berbasis data mining, karena memberikan gambaran nyata mengenai pola penjualan yang telah terjadi.

Selain itu, penelitian ini juga menggunakan kuesioner sebagai metode pengumpulan data pendukung. Kuesioner diarahkan kepada pihak internal perusahaan, seperti manajer maupun staf penjualan, untuk memperoleh wawasan tambahan terkait faktor-faktor yang memengaruhi kinerja penjualan. Pertanyaan dalam kuesioner disusun menggunakan skala Likert, sehingga responden dapat memberikan penilaian secara kuantitatif terhadap berbagai aspek, misalnya persepsi mengenai efektivitas promosi, kepuasan pelanggan, atau faktor internal yang berkontribusi pada peningkatan penjualan. Informasi ini berguna untuk melengkapi data transaksi dan memberikan konteks yang lebih mendalam dalam analisis.

Selain data primer, penelitian ini juga memanfaatkan studi pustaka untuk memperkuat landasan teori dan metodologi. Sekaran dan Bougie yang menyatakan bahwa penentuan sampel yang tepat merupakan kunci dalam menghasilkan temuan penelitian yang valid dan dapat diandalkan [17]. Studi pustaka dilakukan dengan mengkaji berbagai sumber ilmiah yang relevan, seperti buku teks, jurnal akademik, artikel penelitian terdahulu, dan literatur yang membahas konsep data mining, algoritma Decision Tree, serta aplikasi analisis data dalam konteks penjualan F&B. Informasi dari studi pustaka membantu menegaskan kerangka teori, membandingkan temuan dengan penelitian sebelumnya, serta memastikan metode yang digunakan sesuai dengan praktik terbaik dalam bidang ini.

Dengan kombinasi ketiga teknik pengumpulan data ini, penelitian mampu menghadirkan analisis yang komprehensif, menggabungkan data historis yang objektif, persepsi internal yang kontekstual, dan landasan teori yang kuat untuk mendukung validitas serta reliabilitas hasil penelitian.

2.4 Teknik Analisis Data

Dalam penelitian ini, teknik analisis data yang diterapkan mencakup dua pendekatan utama, yaitu analisis deskriptif dan analisis berbasis data mining [18]. Analisis deskriptif digunakan untuk memberikan gambaran awal mengenai karakteristik data penjualan, termasuk distribusi nilai pada masing-masing variabel penelitian, pola umum transaksi, serta kecenderungan penjualan berdasarkan atribut tertentu. Dengan analisis ini, peneliti dapat memahami struktur dan sebaran data sebelum melakukan pemodelan yang lebih kompleks.

Selanjutnya, analisis utama dilakukan dengan memanfaatkan algoritma Decision Tree, yang melalui beberapa tahapan sistematis. Tahapan pertama adalah preprocessing data, di mana data mentah diolah agar bersih, konsisten, dan dalam format yang sesuai untuk pemodelan, termasuk penanganan missing value, penghapusan duplikasi, serta transformasi data ke bentuk kategorikal atau numerik sesuai kebutuhan algoritma. Tahap berikutnya adalah pembentukan model (*modeling*), di mana Decision Tree membangun struktur pohon keputusan berdasarkan atribut-atribut yang tersedia, dengan tujuan mengklasifikasikan tingkat penjualan ke dalam kategori tertentu.

Setelah model terbentuk, dilakukan evaluasi model menggunakan metrik evaluasi seperti akurasi dan confusion matrix. Evaluasi ini berfungsi untuk mengukur sejauh mana model mampu mengklasifikasikan data secara benar, sekaligus mengidentifikasi jenis kesalahan prediksi yang terjadi. Hasil evaluasi memberikan informasi penting mengenai keandalan dan performa model dalam konteks klasifikasi penjualan.

Tahap terakhir adalah interpretasi hasil analisis, di mana pola dan hubungan antarvariabel yang ditemukan model dijabarkan dalam bentuk aturan keputusan (*decision rules*). Aturan-aturan ini berupa logika “jika–maka” yang menjelaskan kondisi tertentu yang mengarah pada kategori penjualan tertentu. Dengan demikian, hasil analisis tidak hanya berfungsi sebagai prediksi, tetapi juga menjadi dasar yang kuat untuk pengambilan keputusan strategis, seperti penyesuaian harga, optimasi promosi, atau penentuan strategi pemasaran yang lebih tepat sasaran.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil pengolahan data serta pembahasan dari penerapan algoritma Decision Tree pada data penjualan perusahaan *Food and Beverage* (F&B). Analisis dilakukan secara bertahap mulai dari proses *preprocessing* data, pembentukan model klasifikasi, hingga evaluasi kinerja model. Hasil yang diperoleh kemudian diinterpretasikan untuk menghasilkan informasi yang relevan dalam mendukung pengambilan keputusan strategis berbasis data (*data-driven decision making*).

3.1 Hasil Pengolahan Data

3.1.1 Gambaran Umum Data Penelitian

Objek dalam penelitian ini adalah sebuah perusahaan yang beroperasi di sektor Food and Beverage (F&B), yang secara konsisten menjalankan aktivitas penjualan produk makanan dan minuman sebagai bagian utama dari kegiatan bisnisnya. Dalam proses operasional sehari-hari, perusahaan tersebut secara otomatis menghasilkan data transaksi penjualan yang tersimpan dalam sistem informasi berbasis digital. Data ini mencerminkan seluruh aktivitas penjualan yang terjadi dalam periode tertentu dan menjadi sumber informasi penting bagi analisis bisnis.

Data yang digunakan dalam penelitian ini berupa data historis transaksi penjualan, yaitu data yang merekam aktivitas penjualan yang telah berlangsung sebelumnya. Dataset tersebut terdiri atas beberapa atribut utama yang relevan dengan proses analisis, di antaranya waktu terjadinya transaksi, jenis atau nama produk yang dijual, harga produk, jumlah unit yang terjual, serta metode pembayaran yang digunakan oleh pelanggan. Seluruh data tersebut disajikan dalam bentuk dataset terstruktur, sehingga memudahkan dalam proses pengolahan dan analisis menggunakan teknik data mining.

Secara umum, data penelitian memiliki karakteristik sebagai berikut

- a. bersifat historis karena berasal dari transaksi yang telah terjadi.
- b. bersifat kuantitatif sehingga dapat diolah secara numerik maupun kategorikal.
- c. memiliki banyak variabel (*multivariat*).
- d. memiliki volume data yang cukup besar.

Karakteristik-karakteristik tersebut menunjukkan bahwa dataset yang digunakan telah memenuhi kriteria yang diperlukan untuk penerapan metode data mining, khususnya algoritma Decision Tree. Struktur data yang jelas, jumlah data yang memadai, serta keberagaman variabel yang terkandung di dalamnya memungkinkan proses analisis dilakukan secara optimal. Selain itu, data tersebut juga mengandung potensi informasi yang dapat digali lebih dalam untuk mengidentifikasi pola, hubungan antarvariabel, serta faktor-faktor yang berpengaruh terhadap kinerja penjualan. Dengan demikian, hasil analisis yang diperoleh diharapkan dapat memberikan kontribusi yang signifikan dalam mendukung pengambilan keputusan berbasis data secara lebih efektif dan terarah.

3.1.2 Seleksi dan Pembersihan Data

Tahap awal dalam proses pengolahan data pada penelitian ini adalah melakukan seleksi terhadap atribut-atribut yang dianggap relevan dengan tujuan analisis. Tahapan ini menjadi penting karena tidak semua data yang terdapat dalam sistem transaksi memiliki kontribusi signifikan terhadap pembentukan pola penjualan. Oleh karena itu, diperlukan proses penyaringan untuk memastikan bahwa hanya atribut yang benar-benar memiliki keterkaitan dengan variabel yang diteliti yang akan digunakan dalam proses selanjutnya.

Dalam praktiknya, data transaksi penjualan biasanya mengandung berbagai jenis atribut, baik yang bersifat informatif maupun administratif. Atribut seperti nomor transaksi, kode invoice, atau identitas kasir, meskipun penting untuk keperluan pencatatan operasional, tidak memiliki pengaruh langsung terhadap analisis pola penjualan. Atribut-atribut tersebut cenderung bersifat unik (*unique identifier*) dan tidak memberikan nilai tambah dalam proses klasifikasi atau prediksi, sehingga tidak relevan untuk dimasukkan ke dalam model analisis.

Oleh karena itu, tahap awal ini difokuskan pada proses identifikasi dan pemilihan atribut yang memiliki potensi untuk memengaruhi kinerja penjualan. Pemilihan atribut dilakukan dengan mempertimbangkan hubungan logis antara variabel dengan tujuan penelitian, serta potensi kontribusinya dalam membentuk pola data. Dengan melakukan seleksi atribut yang tepat, proses analisis menjadi lebih efisien, mengurangi kompleksitas model, serta meningkatkan akurasi hasil yang diperoleh.

Berdasarkan pertimbangan tersebut, atribut-atribut yang dipilih dalam penelitian ini merupakan atribut yang secara langsung maupun tidak langsung berkaitan dengan aktivitas penjualan, yaitu:

- a. Hari penjualan
- b. Jenis produk
- c. Harga produk
- d. Jumlah terjual
- e. Metode pembayaran
- f. Status penjualan

Setelah proses seleksi, dilakukan tahap pembersihan data (*data cleaning*). Tahap ini bertujuan untuk meningkatkan kualitas dataset agar terbebas dari kesalahan yang dapat mempengaruhi hasil analisis meliputi:

- a. penanganan *missing value*.
- b. penghapusan data duplikat.
- c. standarisasi penulisan data.

Proses ini menghasilkan dataset yang lebih konsisten dan siap digunakan dalam analisis

3.1.3 Transformasi Data

Transformasi data merupakan tahapan penting yang dilakukan untuk menyesuaikan bentuk dan struktur data agar sesuai dengan kebutuhan algoritma Decision Tree, sehingga proses analisis dan klasifikasi dapat berjalan secara optimal. Pada tahap ini, data mentah yang sebelumnya masih dalam kondisi belum terorganisir dengan baik diolah menjadi data yang lebih sistematis, konsisten, dan siap digunakan dalam proses pemodelan. Transformasi ini juga bertujuan untuk memastikan bahwa setiap atribut dapat direpresentasikan dalam format yang sesuai, khususnya dalam bentuk kategorikal yang lebih mudah diproses oleh algoritma.

Selain itu, proses transformasi data membantu dalam menyederhanakan kompleksitas data tanpa menghilangkan informasi penting yang terkandung di dalamnya. Dengan demikian, pola-pola yang ada dalam data dapat lebih mudah dikenali oleh model, sehingga hasil analisis yang diperoleh menjadi lebih akurat dan dapat diinterpretasikan dengan lebih jelas.

Adapun tahapan transformasi data yang dilakukan dalam penelitian ini meliputi beberapa proses utama sebagai berikut:

- a. pengelompokan hari menjadi *weekday* dan *weekend*.

Data tanggal transaksi dikonversi menjadi kategori hari, kemudian dikelompokkan menjadi dua kategori utama, yaitu *weekday* (hari kerja) dan *weekend* (akhir pekan). Transformasi ini dilakukan untuk mengidentifikasi pola penjualan berdasarkan waktu operasional.

- b. kategorisasi harga produk:

Pada tahap ini, variabel harga produk dilakukan pengelompokan ke dalam beberapa kategori berdasarkan rentang nilai tertentu, yaitu murah, sedang, dan mahal. Pengelompokan ini bertujuan untuk menyederhanakan variasi harga yang sebelumnya bersifat numerik menjadi bentuk kategorikal, sehingga lebih mudah dipahami oleh algoritma Decision Tree dalam mengenali pola hubungan antara harga dan tingkat penjualan.

Tabel 2. Kategori Harga

Rentang Harga	Kategori
< Rp15.000	Murah
Rp15.000 – Rp30.000	Sedang
> Rp30.000	Mahal

Pengelompokan ini dilakukan untuk menyederhanakan variasi harga dan membantu algoritma dalam menentukan pola hubungan terhadap kinerja penjualan.

- c. kategorisasi jumlah penjualan:

Pada tahap ini, variabel jumlah penjualan juga dikonversi menjadi beberapa kategori berdasarkan tingkat kuantitas penjualan yang terjadi. Proses ini dilakukan untuk mengubah data numerik menjadi bentuk kategorikal sehingga lebih sederhana dalam proses analisis dan klasifikasi. Dengan adanya pengelompokan ini, model dapat lebih mudah mengidentifikasi pola penjualan.

Tabel 3. Kategori Jumlah Penjualan

Jumlah Terjual	Kategori
1-10	Rendah
11-30	Sedang
> 30	Tinggi

Kategorisasi ini bertujuan untuk mengelompokkan tingkat penjualan sehingga dapat digunakan sebagai indikator dalam menentukan keberhasilan penjualan.

d. pembentukan variabel target:

Tabel 4. Status Penjualan

Kondisi	Status
\geq Target	Tercapai
$<$ Target	Tidak Tercapai

Variabel ini menjadi label klasifikasi pada model Decision Tree.

3.1.4 Dataset Hasil Preprocessing

Dataset akhir yang digunakan dalam penelitian ini telah disusun secara terstruktur untuk memudahkan proses analisis dan pemodelan menggunakan algoritma Decision Tree. Penyusunan dataset ini memastikan setiap atribut memiliki format yang konsisten, lengkap, dan siap diolah oleh algoritma. Struktur dataset ini dirancang agar memuat semua informasi yang relevan terkait transaksi penjualan, sehingga dapat merepresentasikan kondisi bisnis secara menyeluruh. Secara umum, dataset akhir tersusun dengan kolom-kolom atau atribut sebagai berikut:

Tabel 5. Contoh Dataset

Hari	Produk	Harga	Jumlah	Pembayaran	Status
Weekend	Minuman	Murah	Tinggi	Non Tunai	Tercapai
Weekday	Makanan	Sedang	Rendah	Tunai	Tidak Tercapai

Dataset ini digunakan sebagai input dalam proses pemodelan.

3.2 Implementasi dan Evaluasi Model

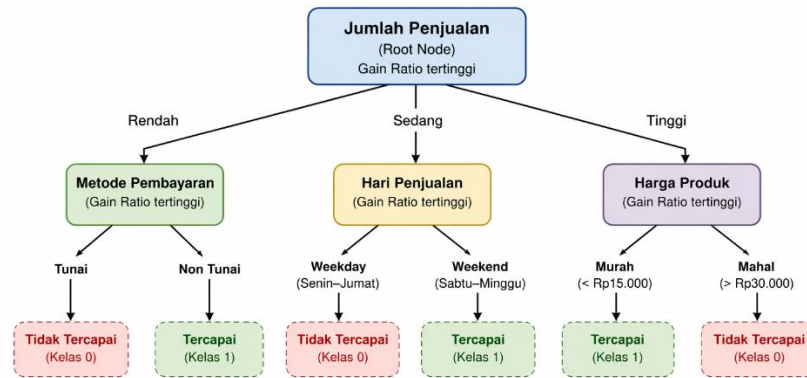
3.2.1 Penerapan Algoritma Decision Tree

Dataset yang telah melalui seluruh tahapan praproses dan transformasi selanjutnya dianalisis dengan menggunakan algoritma Decision Tree. Pada tahap ini, data dimanfaatkan untuk membangun model klasifikasi yang mampu mengidentifikasi pola hubungan antar variabel dalam menentukan kategori atau status penjualan. Hasil dari proses ini direpresentasikan dalam bentuk struktur pohon keputusan (decision tree), yang menggambarkan alur pengambilan keputusan berdasarkan atribut-atribut yang paling berpengaruh.

Struktur pohon yang dihasilkan menunjukkan bagaimana setiap variabel berperan dalam proses klasifikasi, dimulai dari atribut yang memiliki pengaruh paling besar sebagai akar (root node), hingga atribut pendukung lainnya pada cabang-cabang berikutnya. Setiap percabangan dalam pohon merepresentasikan kondisi atau nilai tertentu dari suatu atribut, yang kemudian mengarah pada keputusan akhir berupa klasifikasi tingkat penjualan. Dengan demikian, model ini tidak hanya berfungsi sebagai alat prediksi, tetapi juga sebagai sarana untuk memahami hubungan logis antar variabel dalam data.

Dari hasil pemodelan tersebut, diperoleh sejumlah aturan keputusan (decision rules) yang merupakan hasil turunan dari struktur pohon keputusan. Aturan-aturan ini berbentuk logika “jika–maka” (if–then rules) yang menjelaskan kondisi tertentu yang mengarah pada suatu klasifikasi penjualan. Setiap aturan mencerminkan pola yang ditemukan dalam data, sehingga dapat digunakan sebagai dasar dalam memahami faktor-faktor yang memengaruhi kinerja penjualan serta sebagai acuan dalam pengambilan keputusan strategis.

Selain itu, hasil analisis ini juga memberikan gambaran yang lebih mudah dipahami oleh pihak non-teknis karena struktur pohon keputusan bersifat visual dan intuitif. Hal ini memudahkan interpretasi hasil model tanpa harus memahami perhitungan matematis yang kompleks.



Gambar 2. Struktur Pohon Keputusan

Adapun beberapa aturan keputusan (decision rules) yang dihasilkan dari model tersebut adalah sebagai berikut:

- jika harga murah dan jumlah penjualan tinggi, maka status penjualan tercapai.
- jika harga mahal dan jumlah penjualan rendah, maka status penjualan tidak tercapai.
- jika hari penjualan weekend dan jumlah penjualan sedang atau tinggi, maka cenderung tercapai.
- jika metode pembayaran tunai dengan jumlah rendah, maka cenderung tidak tercapai.

Aturan-aturan ini menunjukkan bagaimana kombinasi atribut memengaruhi hasil klasifikasi dan dapat digunakan sebagai dasar pengambilan keputusan. Dalam algoritma C4.5, pemilihan atribut pada setiap node dilakukan berdasarkan nilai Gain Ratio tertinggi, yaitu ukuran yang digunakan untuk menentukan atribut terbaik dalam membagi data. Atribut dengan nilai Gain Ratio tertinggi akan dipilih sebagai node utama karena mampu mengurangi tingkat ketidakpastian (entropy) secara maksimal.

3.2.2 Evaluasi Model

Evaluasi kinerja model dilakukan dengan memanfaatkan **confusion matrix**, yaitu salah satu metode standar dalam pengukuran performa model klasifikasi. Confusion matrix memberikan gambaran yang rinci mengenai sejauh mana prediksi yang dilakukan oleh model sesuai dengan data aktual. Dengan menggunakan matriks ini, kita dapat menilai kemampuan model dalam mengklasifikasikan setiap kategori penjualan dengan lebih komprehensif, bukan hanya sekadar melihat akurasi keseluruhan.

Tabel 6. Confusion Matrix

Aktual	Tercapai	Tidak Tercapai
Tercapai	45	5
Tidak	7	43

Berdasarkan tabel tersebut, diperoleh hasil sebagai berikut:

- True Positive (TP) = 45
- False Negative (FN) = 5
- False Positive (FP) = 7
- True Negative (TN) = 43

Nilai akurasi model dihitung sebagai berikut:

$$\text{Akurasi} = (TP + TN) / (TP + TN + FP + FN) = 88\%$$

Selain akurasi, dilakukan perhitungan metrik evaluasi lainnya untuk memberikan gambaran performa model secara lebih komprehensif, yaitu precision, recall, dan F1-score.

- Precision = $45 / (45 + 7) = 86,5\%$
- Recall = $45 / (45 + 5) = 90\%$
- F1-score = $88,2\%$

Nilai recall yang tinggi menunjukkan bahwa model mampu mengidentifikasi data dengan status “tercapai” dengan baik, sedangkan nilai precision menunjukkan bahwa prediksi yang dihasilkan cukup akurat. Nilai F1-score yang seimbang menunjukkan bahwa model memiliki performa yang stabil dalam melakukan klasifikasi.

Secara keseluruhan, model memiliki tingkat kesalahan yang relatif rendah, yaitu sebesar 12% dari total data. Hal ini menunjukkan bahwa model cukup andal dalam membedakan kedua kelas.

Untuk meningkatkan validitas model, teknik seperti *k-fold cross-validation* dapat diterapkan pada penelitian selanjutnya. Metode ini memungkinkan model diuji pada berbagai subset data sehingga hasil evaluasi menjadi lebih stabil dan tidak bergantung pada satu pembagian data saja.

3.2.3 Pembahasan Hasil

Hasil penelitian menunjukkan bahwa penerapan algoritma Decision Tree dengan metode C4.5 mampu mengidentifikasi pola penjualan secara efektif dan sistematis. Hal ini dibuktikan melalui hasil evaluasi model yang menunjukkan tingkat akurasi sebesar 88%, serta didukung oleh nilai precision, recall, dan F1-score yang masing-masing mencapai 86,5%, 90%, dan 88,2%. Nilai-nilai tersebut menunjukkan bahwa model memiliki performa yang baik dan stabil dalam mengklasifikasikan status penjualan.

Secara teknis, dominasi variabel jumlah penjualan dan harga produk dalam model menunjukkan bahwa kedua atribut tersebut memiliki nilai Gain Ratio tertinggi. Dalam algoritma C4.5, atribut dengan nilai Gain Ratio tertinggi akan dipilih sebagai node utama karena memiliki kemampuan terbaik dalam mengurangi tingkat ketidakpastian (entropy). Hal ini menjelaskan mengapa variabel jumlah penjualan menjadi root node dalam struktur pohon keputusan, diikuti oleh atribut harga, waktu penjualan, dan metode pembayaran sebagai cabang keputusan.

Dari sisi bisnis, hasil ini menunjukkan bahwa jumlah penjualan merupakan indikator utama dalam menentukan keberhasilan penjualan, sedangkan harga produk berpengaruh terhadap daya beli konsumen. Produk dengan harga yang lebih rendah cenderung memiliki tingkat penjualan yang lebih tinggi, sehingga lebih mudah mencapai target penjualan. Temuan ini relevan dengan konsep perilaku konsumen yang menyatakan bahwa harga merupakan salah satu faktor utama dalam keputusan pembelian.

Selain itu, faktor waktu juga berpengaruh terhadap pola penjualan. Transaksi pada periode weekend cenderung menghasilkan penjualan yang lebih tinggi dibandingkan weekday, yang menunjukkan adanya pola perilaku konsumen berdasarkan waktu. Hal ini dapat dimanfaatkan oleh perusahaan dalam menyusun strategi promosi dan pengelolaan stok secara lebih optimal pada periode tertentu.

Metode pembayaran juga memberikan kontribusi terhadap hasil klasifikasi. Penggunaan metode pembayaran non tunai cenderung berkorelasi dengan tingkat penjualan yang lebih tinggi, yang menunjukkan bahwa kemudahan dan fleksibilitas dalam transaksi dapat meningkatkan minat beli konsumen. Fenomena ini sejalan dengan perkembangan sistem pembayaran digital yang semakin berkembang di sektor bisnis F&B.

Jika dibandingkan dengan metode klasifikasi lain seperti Naïve Bayes dan Random Forest, algoritma Decision Tree memiliki keunggulan utama dalam hal interpretasi model. Meskipun metode seperti Random Forest berpotensi menghasilkan akurasi yang lebih tinggi, Decision Tree lebih mudah dipahami karena menghasilkan aturan keputusan yang eksplisit dan transparan. Hal ini menjadi keunggulan penting dalam konteks bisnis, di mana pihak manajemen membutuhkan model yang tidak hanya akurat tetapi juga mudah diinterpretasikan.

Hasil penelitian ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa algoritma Decision Tree efektif dalam mengidentifikasi pola penjualan serta menghasilkan model yang mudah diinterpretasikan. Namun demikian, algoritma ini juga memiliki beberapa keterbatasan, seperti potensi overfitting serta sensitivitas terhadap perubahan data. Selain itu, proses kategorisasi data dapat menyebabkan hilangnya informasi detail dari data numerik. Implikasi praktis dari penelitian ini menunjukkan bahwa perusahaan dapat:

- a. menentukan strategi harga yang lebih kompetitif,
- b. mengoptimalkan penjualan pada waktu tertentu,
- c. meningkatkan penggunaan metode pembayaran digital.

Secara keseluruhan, penelitian ini menunjukkan bahwa pemanfaatan data mining dengan algoritma Decision Tree mampu memberikan insight yang bernilai dalam mendukung pengambilan keputusan berbasis data. Hasil yang diperoleh tidak hanya bersifat deskriptif, tetapi juga memiliki nilai prediktif yang dapat digunakan dalam perencanaan strategi bisnis yang lebih efektif dan terarah

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan metode data mining menggunakan algoritma Decision Tree (C4.5) terbukti efektif dalam menganalisis data penjualan pada perusahaan *Food and Beverage* (F&B). Proses analisis dilakukan melalui beberapa tahapan utama, yaitu preprocessing data, transformasi data, pembentukan model klasifikasi, serta evaluasi kinerja model. Dataset yang digunakan telah melalui proses seleksi atribut, pembersihan data, dan kategorisasi sehingga menghasilkan data yang berkualitas dan siap untuk dianalisis.

Model Decision Tree yang dibangun mampu mengklasifikasikan status penjualan dengan baik berdasarkan beberapa variabel penting, yaitu hari penjualan, jenis produk, kategori harga, jumlah penjualan, dan metode

pembayaran. Hasil pemodelan menghasilkan struktur pohon keputusan yang menunjukkan bahwa variabel jumlah penjualan menjadi faktor paling dominan (root node) dalam menentukan keberhasilan penjualan. Selain itu, atribut harga produk, waktu penjualan (weekday dan weekend), serta metode pembayaran juga berkontribusi dalam membentuk pola klasifikasi.

Berdasarkan hasil evaluasi menggunakan confusion matrix, model menghasilkan tingkat akurasi sebesar 88%, yang menunjukkan bahwa model memiliki performa yang baik dalam mengklasifikasikan data penjualan. Nilai akurasi ini mengindikasikan bahwa sebagian besar prediksi yang dihasilkan oleh model telah sesuai dengan kondisi aktual. Selain itu, keunggulan utama dari algoritma Decision Tree adalah kemampuannya dalam menghasilkan aturan keputusan (decision rules) yang sederhana, transparan, dan mudah dipahami. Hal ini sangat bermanfaat dalam konteks bisnis karena hasil analisis dapat langsung digunakan oleh pihak manajemen sebagai dasar pengambilan keputusan strategis.

Hasil analisis juga menunjukkan bahwa pola penjualan dipengaruhi oleh kombinasi beberapa faktor, terutama jumlah penjualan, kategori harga, dan waktu transaksi. Penjualan cenderung meningkat pada periode weekend serta pada produk dengan harga yang lebih terjangkau. Selain itu, metode pembayaran non tunai juga menunjukkan kecenderungan berasosiasi dengan tingkat penjualan yang lebih tinggi, yang mencerminkan perubahan perilaku konsumen dalam bertransaksi di era digital.

Secara keseluruhan, penelitian ini telah berhasil mencapai tujuan utama, yaitu membangun model klasifikasi berbasis Decision Tree yang mampu memberikan insight terhadap pola penjualan serta mendukung pengambilan keputusan berbasis data (data-driven decision making). Namun demikian, penelitian ini masih memiliki beberapa keterbatasan, di antaranya belum dilakukan perbandingan dengan algoritma klasifikasi lain seperti Random Forest atau Naïve Bayes, serta belum dilakukan validasi model menggunakan teknik seperti cross-validation. Oleh karena itu, penelitian selanjutnya disarankan untuk melakukan pengujian dengan metode lain guna memperoleh perbandingan performa yang lebih komprehensif, serta mengembangkan sistem berbasis aplikasi atau real-time analytics agar hasil penelitian dapat diimplementasikan secara langsung dalam lingkungan bisnis.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan, bantuan, dan kontribusi dalam pelaksanaan penelitian ini.

Ucapan terima kasih secara khusus penulis sampaikan kepada dosen pembimbing yang telah memberikan arahan, bimbingan, serta masukan yang sangat berharga selama proses penyusunan penelitian ini. Penulis juga mengucapkan terima kasih kepada pihak perusahaan Food and Beverage (F&B) yang telah memberikan izin serta menyediakan data yang dibutuhkan dalam penelitian ini.

Selain itu, penulis turut mengucapkan terima kasih kepada seluruh responden dan pihak terkait yang telah bersedia memberikan informasi serta membantu kelancaran proses pengumpulan data. Dukungan dari keluarga dan rekan-rekan juga sangat berarti dalam memberikan motivasi dan semangat selama proses penelitian berlangsung.

Penulis menyadari bahwa penelitian ini tidak akan terselesaikan dengan baik tanpa adanya bantuan dan dukungan dari berbagai pihak tersebut. Semoga hasil penelitian ini dapat memberikan manfaat dan kontribusi bagi pengembangan ilmu pengetahuan serta praktik bisnis, khususnya dalam bidang analisis data penjualan.

REFERENCES

- [1] Badan Pusat Statistik, "Statistik Penyediaan Makanan dan Minuman 2023," 2024.
- [2] E. Purwanto, B. P. C. Utomo, H. Permatasari, and F. Mohd, "Comparative Analysis of Classification Models for Sales Prediction in E-commerce: Decision Tree, Random Forest, SVM, Naive Bayes, and KNN," *Jurnal Teknik Informatika (JUTIF)*, 2026.
- [3] J. Smith and M. Rodriguez, "Decision Tree Algorithm for Classifying Customer Behavior Patterns in Retail Sales Data," *International Journal of Data Science and Analytics*, 2023.
- [4] R. D. Kurniawan, D. N. D. Sukarman, K. W. Rumaropen, and C. B. Allo, "Analisis Komparatif Algoritma Decision Tree dan Random Forest untuk Klasifikasi Penjualan Produk pada Dataset Superstore," *STATMAT: Jurnal Statistika dan Matematika*, 2025.
- [5] R. Beay and F. M. Sarimole, "Application of Decision Tree Method for Sales Prediction at PT. Cipta Naga Semesta (Mayora Group) North Jakarta for 2023," *International Journal Software Engineering and Computer Science*, 2024.
- [6] M. Iqbal, S. Miskiyah, et al., "Perbandingan Metode Decision Tree dan Naive Bayes pada Tingkat Penjualan Minuman Kopi," *Jurnal INSAN: Journal of Information System Management Innovation*, 2024.

- [7] Kurniawan et al. (2025). Analisis Komparatif Algoritma Decision Tree dan Random Forest untuk Klasifikasi Penjualan Produk pada Dataset Superstore.
- [8] J. W. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 4th ed. Thousand Oaks, CA: Sage Publications, 2014.
- [9] Hadir, "Current Trends in Indonesia's Food & Beverage Industry," 2025.
- [10] D. I. Suranda and A. Nugroho, "Klasifikasi Data Penjualan untuk Memprediksi Tingkat Penjualan Produk Menggunakan Metode Decision Tree," *Jurnal Teknik Informasi dan Komputer*, 2025.
- [11] S. Wibisono, "Aplikasi Prediksi Penjualan AC Menggunakan Decision Tree dengan Algoritma C4.5," *Jurnal Misi*, 2020.
- [12] A. Izyuddin and S. Wibisono, "Aplikasi Prediksi Penjualan Menggunakan Decision Tree," *Jurnal Manajemen Informatika & Sistem Informasi*.
- [13] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Morgan Kaufmann, 2012.
- [14] P. Kotler and K. L. Keller, *Marketing Management*, 15th ed. Pearson Education, 2016.
- [15] Kotler, P., & Keller, K. L. (2016). *Marketing Management*. Pearson Education.
- [16] Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta, 2019.
- [17] U. Sekaran and R. Bougie, *Research Methods for Business: A Skill-Building Approach*, 7th ed. Wiley, 2016.
- [18] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis*, 8th ed. Cengage, 2019.