

Algoritma Data Mining Menggunakan Metode Decision Tree Untuk Memprediksi Pola Penjualan Produk Springbed Menggunakan Algoritma C4.5

Donny Sanjaya¹, Amir Saleh^{2,*}, Sri Novida Sari¹, Surizar Rahmi Danur³

¹Jurusan Teknik Komputer dan Informatika, Program Studi Teknologi Rekayasa Perangkat Lunak, Politeknik Negeri Medan, Medan, Indonesia

²Jurusan Teknik Komputer dan Informatika, Program Studi Teknik Komputer, Politeknik Negeri Medan, Medan, Indonesia

³Fakultas Sains dan Teknologi, Program Studi Sistem Informasi, Institut Teknologi dan Bisnis Indonesia, Deli Serdang, Indonesia

Email: ¹d.sanjaya@polmed.ac.id, ^{2,*}amirsalehst1990@gmail.com, ³srinovidasari@gmail.com, ⁴surizar.rdanur@gmail.com

Email Penulis Korespondensi: amirsalehst1990@gmail.com

Abstrak— Permasalahan yang sering terjadi dalam dunia bisnis penjualan springbed adalah sering terjadinya naik turun prediksi, sulitnya mendeteksi pola terhadap hal apa saja yang dapat meningkatkan penjualan dari pembeli membuat pihak pelaku bisnis penjualan springbed sering mengalami kerugian, hal tersebut juga terjadi karena pelaku bisnis tidak mengetahui strategi yang pasti dalam meningkatkan penjualan sehingga perlu dilakukan sebuah prediksi dengan tingkat akurasi yang tinggi, salah satunya dengan bantuan penerapan ilmu komputer data mining dengan menggunakan metode C4.5. Metode C4.5 yang digunakan dalam penelitian ini mampu menghasilkan pohon keputusan yang optimal, dengan kemampuan untuk memilah atribut yang paling relevan dalam prediksi penjualan springbed. Penggunaan algoritma data mining ini diharapkan dapat memberikan wawasan kepada pelaku bisnis springbed dalam mengambil keputusan strategis, seperti pengelolaan stok, perencanaan produksi, dan kampanye pemasaran yang lebih efektif. Langkah-langkah eksperimen dalam penelitian ini meliputi pengumpulan data penjualan springbed. Hasil eksperimen menunjukkan bahwa algoritma Decision Tree dengan metode C4.5 mampu memberikan prediksi penjualan springbed dengan tingkat akurasi yang memadai. Model ini dapat membantu pelaku bisnis penjualan Springbed dalam merencanakan strategi bisnis yang lebih tepat berdasarkan perkiraan permintaan pasar untuk meningkatkan naik turunnya penjualan.

Kata Kunci: Data Mining; Decision Tree; C4.5; Prediksi; Pola Penjualan Springbed

Abstract— Problems that often occur in the world of spring bed sales business are frequent ups and downs in predictions, the difficulty of detecting patterns in what can increase sales from buyers makes spring bed sales business people often experience losses, this also happens because business people don't know the strategy. Certainly in increasing sales, it is necessary to make predictions with a high level of accuracy, one of which is with the help of the application of computer science data mining using the C4.5 method. The C4.5 method used in this research is able to produce an optimal decision tree, with the ability to sort out the most relevant attributes in predicting springbed sales. The use of this data mining algorithm is expected to provide insight to springbed business players in making strategic decisions, such as stock management, production planning and more effective marketing campaigns. The experimental steps in this research include collecting springbed sales data. Experimental results show that the Decision Tree algorithm using the C4.5 method is able to provide spring bed sales predictions with an adequate level of accuracy. This model can help Springbed sales business players in planning more appropriate business strategies based on estimated market demand to increase the ups and downs of sales.

Keywords: Data Mining; Decision Tree; C4.5; Prediction; Springbed Sales Pattern

1. PENDAHULUAN

Persaingan dalam dunia bisnis menuntut para pengusaha untuk menemukan suatu pola yang dapat meningkatkan harga penjualan dan pemasaran produk di perusahaannya, jika hal tersebut tidak dilakukan maka akan menimbulkan banyak kerugian salah satunya tidak jelasnya pengeluaran dan pendapatan yang diterima oleh perusahaan dalam mencapai target dan cita-cita terwujudnya sebuah perusahaan tersebut sehingga banyak perusahaan yang melakukan sebuah prediksi terhadap pola penjualan[1]. Hasil prediksi pola penjualan dapat digunakan perusahaan untuk menghindari resiko terburuk yang akan terjadi di masa depan terhadap pola penjualan dan dapat menerapkan strategi bisnis dalam mempertahankan eksistensi perusahaan tersebut[2].

Data transaksi yang selama ini dikumpulkan oleh perusahaan tentunya membantu dalam menggali informasi yang dibutuhkan perusahaan khususnya dalam mengetahui pola penjualan perusahaan[3]. Pola penjualan adalah strategi atau taktik yang digunakan oleh perusahaan atau individu untuk menjual produk atau layanan mereka kepada pelanggan., pada bidang ini pola penjualan yang sering mengalami naik turun pasar adalah furniture dengan jenis springbed. Dengan diketahuinya pola penjualan springbed maka perusahaan akan lebih mengetahui kebutuhan pasar dan konsumen serta pengeluaran dan pemasukan yang lebih jelas dalam kurun waktu tertentu. Hal tersebut juga membantu dalam mengetahui perkembangan bisnis perusahaan dengan pasti sehingga dapat membantu membentuk strategi baru dalam mengembangkan bisnis perusahaannya[4].

Permasalahan yang sering terjadi pada adalah dalam mengetahui pola penjualan adalah adanya kegiatan operasional terhadap jual beli atau data transaksi sehari-hari yang input secara manual membuat data semakin lama akan semakin bertambah banyak. Jumlah data yang begitu besar justru bisa menjadi masalah bagi perusahaan tersebut karena tidak bisa dimanfaatkan sebaik mungkin untuk mendapatkan informasi yang dibutuhkan perusahaan terhadap pola penjualan springbed. Semakin banyak data, maka perusahaan tersebut semakin memerlukan usaha berupa waktu yang cukup lama,

mengolah data yang sudah rusak, mengeluarkan biaya operasional lebih hanya untuk mengolah data-data agar dapat dijadikan informasi yang berguna di masa depan.

Urgensi penelitian ini dilatarbelakangi oleh permasalahan dalam pengelolaan data transaksi penjualan yang masih dilakukan secara manual, sehingga menyebabkan akumulasi data dalam jumlah besar yang tidak termanfaatkan secara optimal. Semakin bertambahnya data justru menimbulkan kendala bagi perusahaan karena membutuhkan waktu yang lama untuk pengolahan, berpotensi mengalami kerusakan atau inkonsistensi, serta meningkatkan biaya operasional hanya untuk menjadikannya sebagai informasi yang berguna. Kondisi tersebut mengakibatkan perusahaan kesulitan dalam mengetahui pola penjualan produk, khususnya springbed, yang sebenarnya sangat penting untuk mendukung pengambilan keputusan strategis. Oleh karena itu, diperlukan suatu pendekatan yang mampu mengolah data secara efektif dan efisien, yaitu dengan menerapkan teknik data mining menggunakan metode decision tree dengan algoritma C4.5, yang mampu mengklasifikasikan data serta menghasilkan pola keputusan dari data historis penjualan, sehingga diharapkan dapat membantu perusahaan dalam memprediksi pola penjualan dan meningkatkan kualitas pengambilan keputusan di masa depan.

Solusi yang dilakukan agar pola penjualan dapat diketahui dengan mudah, cepat dan akurasi tinggi, atau hasil yang maksimal yaitu dapat mengolah data transaksi yang sangat banyak pada penjualan springbed menggunakan pemanfaatan sebuah sistem komputer yang dirancang khusus dalam menggali data yang sangat banyak untuk menghasilkan informasi yang dibutuhkan yaitu menggunakan penerapan ilmu sistem data mining.

Data mining merupakan proses pengolahan dan analisis data dalam jumlah besar untuk menemukan pola, hubungan, dan informasi tersembunyi yang sebelumnya tidak diketahui, sehingga dapat dimanfaatkan sebagai dasar pengambilan keputusan[5]. Dalam konteks bisnis, data mining menjadi sangat penting karena mampu mengubah data transaksi yang bersifat mentah menjadi informasi yang bernilai strategis[6]. Dengan memanfaatkan teknik dan algoritma tertentu, seperti klasifikasi, clustering, dan asosiasi, data mining dapat membantu perusahaan dalam memahami perilaku konsumen, memprediksi tren penjualan, serta mengidentifikasi peluang dan risiko yang mungkin terjadi. Oleh karena itu, penerapan data mining tidak hanya meningkatkan efisiensi dalam pengelolaan data, tetapi juga memberikan keunggulan kompetitif bagi perusahaan dalam menghadapi persaingan bisnis yang semakin ketat[7].

Pada dasarnya data mining memiliki penerapan metode prediksi yang sudah banyak dimanfaatkan dan diterapkan oleh perusahaan terkemuka di dunia, salah satu metode prediksi yang dapat digunakan adalah metode decision tree C4.5 dan untuk hasil yang lebih cepat dan akurat data yang telah terkumpul dianalisa menggunakan bantuan teknologi komputer yang mempermudah melihat hasil akurasi menggunakan metode C4.5[8], [9].

Algoritma C4.5 merupakan salah satu metode dalam teknik data mining yang digunakan untuk membentuk model klasifikasi dalam bentuk pohon keputusan (decision tree)[10]. Algoritma ini merupakan pengembangan dari algoritma ID3, dengan keunggulan utama dalam menangani data yang memiliki nilai kontinu, data yang tidak lengkap (missing value), serta kemampuan melakukan proses pruning untuk mengurangi overfitting[11]. Prinsip kerja C4.5 adalah memilih atribut terbaik sebagai akar pohon berdasarkan nilai gain ratio tertinggi, yaitu hasil perbaikan dari information gain yang mempertimbangkan pembagian atribut secara lebih proporsional. Selanjutnya, data akan dibagi secara rekursif ke dalam cabang-cabang hingga menghasilkan keputusan atau kelas tertentu. Dengan struktur yang berbentuk pohon, algoritma C4.5 mudah dipahami dan diinterpretasikan, sehingga banyak digunakan dalam berbagai bidang, seperti prediksi penjualan, analisis risiko, dan sistem pendukung keputusan[12].

Beberapa penelitian pendukung yang telah dilakukan dengan hasil bahwasannya untuk menganalisis dan mengklasifikasikan data penjualan rokok menggunakan algoritma C4.5. Data diolah melalui tahap preprocessing untuk mengatasi duplikasi dan inkonsistensi. Hasil penelitian menunjukkan bahwa algoritma C4.5 mampu menghasilkan model pohon keputusan dengan tingkat akurasi tinggi sebesar 96,73%, sehingga dapat digunakan untuk menentukan strategi penjualan yang lebih optimal[13]. Penelitian ini mengklasifikasikan produk skincare terlaris berdasarkan atribut seperti kategori dan harga menggunakan algoritma C4.5. Hasil menunjukkan bahwa kedua atribut tersebut menjadi faktor utama dalam menentukan produk terlaris, dengan tingkat akurasi sebesar 68,75%. Penelitian ini juga menekankan pentingnya optimasi model untuk meningkatkan hasil prediksi penjualan[14].

Juga penelitian menggunakan algoritma C4.5 untuk memprediksi penjualan spanduk berdasarkan data historis. Metode decision tree mampu mengidentifikasi produk yang paling sering dipesan serta membantu dalam pengambilan keputusan terkait produksi dan pemasaran. Hasil penelitian menunjukkan bahwa penggunaan data mining dapat meningkatkan akurasi informasi penjualan dan mempermudah analisis data[15]. Serta membandingkan algoritma C4.5 dan Random Forest dalam menganalisis faktor yang memengaruhi penjualan produk serum di platform e-commerce. Hasil menunjukkan bahwa algoritma C4.5 mampu menghasilkan aturan klasifikasi dalam bentuk pohon keputusan, sedangkan Random Forest memberikan informasi tingkat kepentingan variabel. Variabel seperti jumlah pengikut dan jumlah produk menjadi faktor dominan dalam penjualan[16].

Berdasarkan penelitian terdahulu yang telah dipaparkan, terlihat bahwa penerapan algoritma C4.5 dalam analisis data mining telah banyak digunakan untuk klasifikasi dan prediksi penjualan pada berbagai jenis produk, seperti rokok, skincare, spanduk, dan produk e-commerce, dengan hasil yang menunjukkan tingkat akurasi yang cukup baik serta kemampuan dalam menghasilkan pola keputusan yang mudah dipahami. Namun demikian, terdapat beberapa celah penelitian (research gap) yang masih dapat dikembangkan, di antaranya adalah belum adanya penelitian yang secara spesifik mengkaji pola penjualan pada produk springbed, sehingga karakteristik data dan perilaku konsumen pada sektor ini belum banyak dianalisis. Selain itu, sebagian penelitian masih berfokus pada tingkat akurasi model tanpa

mengeksplorasi pemanfaatan hasil pohon keputusan sebagai dasar strategi bisnis yang lebih komprehensif. Di sisi lain, pengolahan data pada penelitian sebelumnya umumnya menggunakan data yang sudah terstruktur dengan baik, sehingga belum banyak membahas kondisi nyata di lapangan yang melibatkan data transaksi manual yang cenderung tidak konsisten dan memiliki potensi kesalahan. Oleh karena itu, penelitian ini hadir untuk mengisi kesenjangan tersebut dengan menerapkan algoritma C4.5 dalam menganalisis data penjualan springbed yang bersumber dari data transaksi nyata, serta menghasilkan pola penjualan yang dapat digunakan sebagai dasar dalam pengambilan keputusan strategis perusahaan.

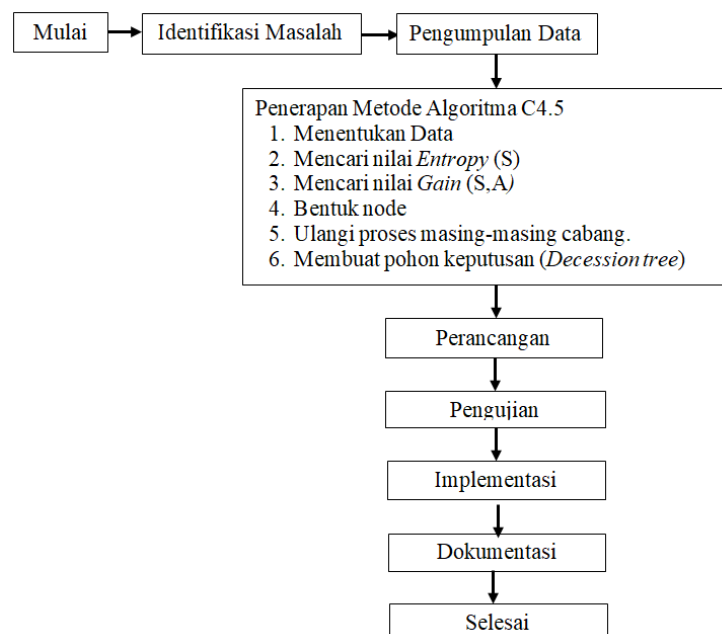
Penelitian ini bertujuan untuk menerapkan teknik data mining dengan metode *decision tree* menggunakan algoritma C4.5 dalam menganalisis data historis penjualan produk springbed guna mengidentifikasi pola penjualan yang terbentuk. Selain itu, penelitian ini juga bertujuan untuk mentransformasikan data transaksi yang sebelumnya dikelola secara manual menjadi informasi yang terstruktur, akurat, dan bernilai guna. Lebih lanjut, penelitian ini diarahkan untuk membangun model klasifikasi yang mampu memprediksi kecenderungan penjualan berdasarkan atribut-atribut yang relevan, serta mengidentifikasi faktor-faktor yang berpengaruh terhadap tingkat penjualan sebagai dasar dalam mendukung proses pengambilan keputusan yang lebih efektif dan berbasis data.

Hasil yang diharapkan dari penelitian ini adalah terbentuknya model klasifikasi dalam bentuk pohon keputusan berbasis algoritma C4.5 yang mampu merepresentasikan pola penjualan produk springbed secara sistematis dan mudah diinterpretasikan. Selain itu, diharapkan model yang dihasilkan memiliki tingkat akurasi yang memadai dalam memprediksi pola penjualan, sehingga dapat dimanfaatkan sebagai alat bantu dalam perencanaan strategi bisnis, seperti pengelolaan persediaan, penentuan prioritas produk, dan pengembangan strategi pemasaran. Penelitian ini juga diharapkan dapat meningkatkan efisiensi dalam pengolahan data transaksi serta memberikan kontribusi ilmiah dalam penerapan teknik data mining untuk mendukung pengambilan keputusan berbasis data pada sektor penjualan.

2. METODOLOGI PENELITIAN

2.1 Kerangka Kerja Penelitian

Kerangka kerja merupakan struktur penelitian yang ditetapkan untuk membantu penulis dalam mencapai target penelitian, tujuan adanya kerangka kerja penelitian untuk mempermudah dalam melakukan tahapan-tahapan yang dilakukan pada saat melakukan penelitian, dengan tahapan-tahapan yang telah ditetapkan maka peneliti akan lebih terstruktur dan berada pada pedoman kerja agar terhindar dari beberapa kegiatan yang tidak diperlukan. Adapun tahapan penelitian yang dilakukan dalam prediksi omset penjualan springbed bawah ini:



Gambar 1. Kerangka Penelitian

Identifikasi masalah dalam penelitian ini difokuskan pada berbagai kendala yang berkaitan dengan penjualan serta omzet penjualan springbed. Permasalahan yang diangkat mencakup ketidakstabilan penjualan, kesulitan dalam memprediksi permintaan pasar, serta kurangnya pemanfaatan data historis sebagai dasar pengambilan keputusan. Dengan mengidentifikasi masalah secara sistematis, peneliti dapat menentukan arah penelitian yang lebih terstruktur dan relevan dengan kondisi nyata di lapangan.

Tahap selanjutnya adalah pengumpulan data, yang dilakukan melalui dua pendekatan utama, yaitu studi literatur dan studi lapangan. Studi literatur dilakukan dengan mengkaji berbagai sumber seperti jurnal ilmiah, buku referensi, serta website yang relevan guna memperkuat landasan teoritis penelitian. Sementara itu, studi lapangan dilakukan dengan cara mengumpulkan data secara langsung dari lokasi penelitian untuk memperoleh informasi yang lebih akurat dan sesuai dengan kondisi aktual. Kombinasi kedua metode ini diharapkan mampu menghasilkan data yang valid dan komprehensif. Pada tahap analisis, peneliti menerapkan algoritma C4.5 untuk mengidentifikasi pola penjualan springbed. Analisis ini bertujuan untuk memahami bagaimana proses klasifikasi data dilakukan, serta bagaimana algoritma tersebut dapat digunakan dalam memprediksi tren penjualan berdasarkan atribut-atribut yang tersedia. Dengan pendekatan ini, diharapkan dapat diperoleh model yang mampu memberikan hasil prediksi yang optimal.

Tahap perancangan dilakukan dengan mengembangkan sebuah sistem atau aplikasi prediksi menggunakan algoritma C4.5 yang diimplementasikan melalui perangkat lunak RapidMiner. Pada tahap ini, peneliti merancang alur kerja sistem, menentukan variabel yang digunakan, serta menguji struktur model yang dibangun untuk mengetahui tingkat akurasi metode yang diterapkan.

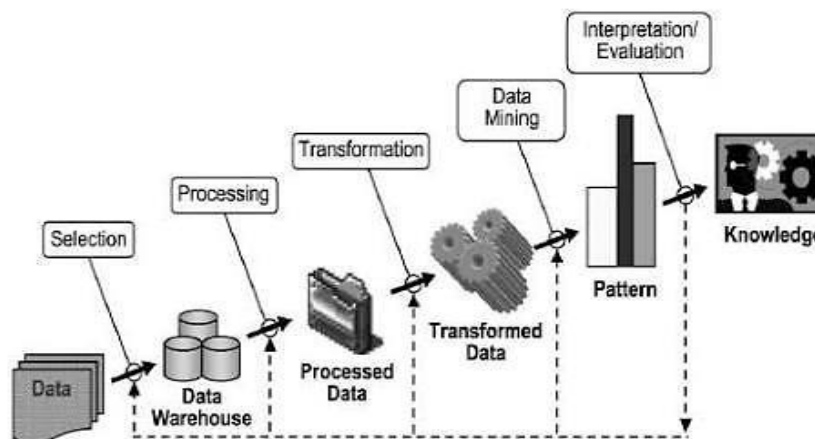
Selanjutnya, dilakukan tahap pengujian untuk memastikan bahwa sistem yang telah dirancang dapat berjalan dengan baik. Pengujian ini bertujuan untuk mengevaluasi kinerja model, mengidentifikasi kemungkinan kesalahan, serta memastikan bahwa hasil prediksi yang dihasilkan sesuai dengan harapan dan kebutuhan penelitian.

Tahap implementasi merupakan proses penerapan sistem yang telah dirancang ke dalam kondisi nyata. Pada tahap ini, peneliti menjalankan langkah-langkah yang telah disusun sebelumnya guna mencapai tujuan penelitian, yaitu menghasilkan model prediksi penjualan yang efektif dan dapat digunakan sebagai alat bantu dalam pengambilan keputusan.

Tahap terakhir adalah dokumentasi, yang mencakup seluruh proses pencatatan dan penyimpanan informasi yang berkaitan dengan penelitian. Dokumentasi dilakukan secara sistematis, mulai dari pengumpulan data, proses analisis, hingga hasil akhir penelitian. Hal ini bertujuan untuk memastikan bahwa seluruh tahapan penelitian terdokumentasi dengan baik sehingga dapat dijadikan referensi untuk penelitian selanjutnya maupun sebagai bukti validitas penelitian yang telah dilakukan.

2.2 Data Mining

Data mining, sering juga disebut knowledge discovery in database (KDD), merupakan kegiatan yang meliputi pengumpulan data, pemakaian data historis untuk menemukan keteraturan pola, atau hubungan dalam set data yang berukuran besar. Bentuk dari data mining bisa dipakai untuk memperbaiki pengambilan keputusan di masa yang akan datang[17]. Data mining bukanlah suatu bidang yang sama sekali baru. Salah satu kesulitan untuk mendefinisikan data mining adalah kenyataan bahwa data mining memberikan banyak aspek dan teknik bidang-bidang ilmu yang sudah permanen terlebih dahulu. data mining memiliki akar panjang dari ilmu seperti ilmu kecerdasan buatan (artificial intelligent), machine learning, statistic, database, dan juga informasi retrieval[18].



Gambar 2. Proses knowledge discovery in database[19]

2.3 Prediksi Pola Penjualan

Prediksi merupakan kegiatan memperkirakan kejadian yang akan berlangsung di masa yang akan mendatang, kegiatan prediksi sudah banyak dilakukan oleh banyak analisis dalam kehidupan dengan tujuan mendapatkan jawaban beberapa pertanyaan dari fenomena-fenomena sebab dan akibat yang akan mendatang[20]. Dengan adanya prediksi dapat membantu pelaku prediksi mendapatkan jawaban dan strategi baru dalam menghadapi beberapa kemungkinan yang terjadi. Prediksi juga dianggap sebagai penyelamat dalam menghadapi resiko buruk yang mungkin akan terjadi di masa depan[21].

Salah satu prediksi yang paling sering dilakukan adalah prediksi kegiatan jual beli dengan tujuan mengetahui naik turunnya penjualan atau pendapatan murni dari hasil penjualan sebuah produk yang akan diperoleh dalam pengambilan

dan penyusunan sebuah perencanaan penjualan. Prediksi dalam penjualan dapat dilakukan dengan mengetahui data dan siklus transaksi yang dilakukan dalam priode waktu tertentu[22]. Tujuan dari mengetahui pola penjualan adalah untuk mengetahui dengan pasti perkembangan dan hasil usaha dari perusahaan dalam mengembangkan bisnisnya[23]. Dengan mempelajari dan memprediksi sebuah pola penjualan akan membantu pelaku usaha dalam menangani beberapa resiko perusahaan dan menetapkan strategi baru perusahaan dalam mencapai tujuan perusahaan lebih terarah dengan cepat dan tepat. Pentingnya melakukan prediksi agar perusahaan dapat menghindari hal-hal yang tidak diinginkan oleh perusahaan di masa depan[24].

2.4 Decision Tree

Decision tree atau pohon keputusan adalah pohon yang digunakan sebagai prosedur penalaran untuk mendapatkan jawaban dari masalah yang dimasukan. Pohon yang dibentuk tidak selalu berupa pohon biner. Decision tree banyak digunakan untuk menyelesaikan kasus penentuan keputusan seperti di bidang medis (diagnosis penyakit pasien), ilmu komputer (struktur data), psikologi (teori pengambilan keputusan)[25].

2.5 Algoritma C4.5

Algoritma C4.5 ialah salah satu metode yang diterapkan untuk membentuk pohon keputusan berdasarkan data pelatihan. Metode ini merupakan salah satu pendekatan yang sangat efektif dan terkenal dalam klasifikasi serta prediksi. Pohon keputusan adalah model prediksi yang mengadopsi kerangka berstruktur hierarkis. Prinsip dari pohon keputusan adalah mengonversi data menjadi pohon keputusan dan aturan-aturan keputusan. Algoritma C4.5 merepresentasikan dari pendekatan ID3[26]. Macam-macam pengembangan yang dilakukan C4.5 adalah dapat mengatasi missing value, continue data dan pruning. Algoritma C4.5 mempunyai input berupa training samples dan samples, training samples berupa data contoh yang akan digunakan untuk membentuk sebuah pohon yang telah diuji kebenarannya, sedangkan untuk samples adalah field-field data yang nantinya akan kita gunakan sebagai kriteria dalam melakukan klasifikasi data. Berikut langkah-langkah dalam menyelesaikan Algoritma C4.5 diantaranya sebagai berikut[27], [28]:

Proses dalam penerapan algoritma C4.5 dimulai dengan menghitung nilai entropy dari setiap kriteria untuk mengukur tingkat ketidakpastian data, kemudian dilanjutkan dengan menghitung nilai gain pada masing-masing atribut guna mengetahui seberapa besar informasi yang diperoleh dalam proses klasifikasi; atribut dengan nilai gain tertinggi selanjutnya dipilih sebagai akar (root) dalam pembentukan pohon keputusan karena dianggap paling efektif dalam memisahkan data, setelah itu dilakukan pembentukan cabang berdasarkan nilai-nilai yang dimiliki oleh atribut tersebut sehingga menghasilkan struktur awal pohon keputusan, dan proses ini terus diulang pada setiap cabang dengan menghitung kembali nilai entropy dan gain pada subset data yang baru hingga seluruh data terklasifikasi secara optimal atau memenuhi kondisi penghentian yang telah ditentukan.

Untuk mencari Rumus pada Algoritma C4.5 pada pohon keputusan adalah sebagai berikut[29], [30]:

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum_{i=1}^n \left| \frac{S_i}{S} \right| * \text{Entropy}(S_i) \quad (1)$$

Di mana:

- S : himpunan kasus
- A : atribut
- n : jumlah partisi atribut A
- |S_i| : jumlah kasus pada partisi ke-i
- |S| : jumlah kasus dalam

$$\text{Entropy}(S) = \text{Entropy}(S) - \sum_{i=1}^n \left| \frac{S_i}{S} \right| * p_i * \log_2 p_i \quad (2)$$

Di mana :

- S : himpunan kasus
- A : atribut
- n : jumlah partisi S
- p_i : proporsi dari S_i terhadap S

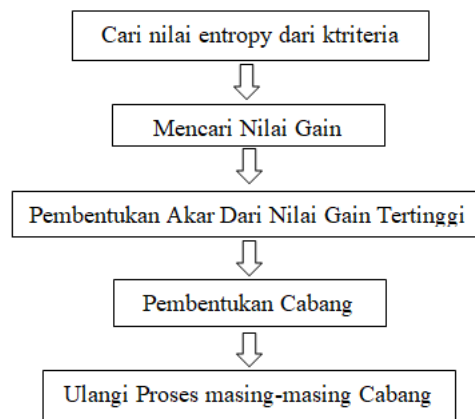
3. HASIL DAN PEMBAHASAN

Sampel data yang digunakan merupakan hasil dan kegiatan yang ada selama pergerakan hasil penjualan springbed dalam kurun waktu satu tahun. Data tersebut menggambarkan kondisi hari seperti adanya peringatan penting pada bulan tertentu, kegiatan promosi yang dilakukan pada bulan-bulan tertentu dengan strategi pemberian diskon, promosi kualitas produk di sosial media, menyebarkan katalog atau brosur produk, persediaan produk dan table keputusan target kondisi hasil yang didapatkan. Adapun sampel dapat dilihat pada tabel 1 berikut ini:

Tabel 1. Sampel Data Hasil Hasil Penjualan

No.	Bulan	Jumlah Barang	Kegiatan Promosi	Kualitas	Hasil Penjualan
1	Januari	350 Set	Pemberian diskon	Tinggi	8 M
2	Februari	160 Set	Brosur Produk	Sedikit	2 M
3	Maret	170 Set	Promosi sosial media	Sedang	2,5 M
4	April	250 Set	Brosur Produk	Sedang	5 M
5	Mei	140 Set	Brosur Produk	Sedikit	900 Jt
6	Juni	200 Set	Promosi sosial media	Sedang	3 M
7	Juli	120 Set	Toko	Sedikit	1 M
8	Agustus	290 Set	Pemberian diskon	Tinggi	6 M
9	September	190 Set	Pemberian diskon	Sedang	5 M
10	Oktober	170 Set	Pemberian diskon	Sedang	4 M
11	November	200 Set	Promosi sosial media	Sedang	3,5 M
12	Desember	400 Set	Promosi sosial media	Tinggi	10 M
13	Januari	400 Set	Toko	Tinggi	10 M
14	Februari	300 Set	Promosi sosial media	Tinggi	8 M
15	Maret	370 Set	Toko	Tinggi	8,5 M

Dilakukannya prediksi terhadap pola penjualan adalah untuk menemukan pola terhadap naik dan turunnya pada penjualan springbed, penerapan metode algoritma C4.5 dilakukan dengan membaca kasus lama sebagai rekam jejak dari beberapa faktor yang mempengaruhi pola penjualan. Setelah ditemukan hasil prediksi maka akan diketahui beberapa faktor yang dapat menyebabkan menurunnya penjualan atau naiknya pola penjualan springbed. Adapun tahapan analisa yang akan dilakukan dengan algoritma C4.5 dapat dilihat pada gambar 3 berikut ini:



Gambar 3. Tahapan Analisa C4.5

3.1 Penerapan Algoritma C4.5

Pada penggunaan Algoritma C4.5 menggunakan data keterangan, data keterangan tersebut akan dihitung berapa kali muncul dalam sampel data, sehingga data angka harus dirubah menjadi data keterangan, perubahan data ini disebut dengan preprocessing data. Untuk mempermudah proses penyelesaian permasalahan menggunakan Algoritma C4.5, berikut ini merupakan tabel data keterangan dari masing-masing nilai angka kriteria.

Tabel 2. Keterangan Kriteria

Kode	Kriteria	Data Angka	Keterangan
1	Status Omset	0-2 Milyar	Sedikit
		2,5-5 Milyar	Sedang
		> 5 Milyar	Tinggi
2	Jumlah Barag	0-160	Sedikit
		161-250	Sedang
		>280	Tinggi

Berdasarkan data keterangan kriteria maka data yang dilakukan preprocessing data terhadap sampel menjadi data baru sebagai berikut ini:

Tabel 3. Sampel Data yang akan diuji

No	Bulan	Jumlah Barang	Kegiatan Promosi	Kualitas	Pola Penjualan
1	Januari	Tinggi	Pemberian diskon	Tinggi	Tinggi
2	Februari	Sedang	Brosur Produk	Sedang	Sedikit
3	Maret	Sedang	Promosi sosial media	Tinggi	Sedikit
4	April	Sedang	Brosur Produk	Tinggi	Sedang
5	Mei	Sedikit	Brosur Produk	Sedikit	Sedikit
6	Juni	Sedang	Promosi sosial media	Sedang	Sedang
7	Juli	Sedikit	Toko	Sedikit	Sedikit
8	Agustus	Sedang	Pemberian diskon	Sedang	Sedang
9	September	Sedang	Pemberian diskon	Sedikit	Sedang
10	Oktober	Sedang	Pemberian diskon	Tinggi	Sedang
11	November	Sedang	Promosi sosial media	Sedang	Sedang
12	Desember	Sedang	Promosi sosial media	Sedikit	Sedang
13	Januari	Banyak	Toko	Tinggi	Tinggi
14	Februari	Sedang	Promosi sosial media	Sedang	Tinggi
15	Maret	Tinggi	Toko	Tinggi	Tinggi

Berdasarkan data di atas maka dalam penyelesaian Algoritma C4.5 dilakukan Pembentukan akar awal.
 Pertama : Pembentukan akar awal atau node awal yaitu dengan mencari nilai Entropy dan nilai Gain.

Tabel 4. Node 1 Pembentukan Akar Awal

Node	Keterangan	Jumlah Kasus	Rendah	Sedang	Tinggi
1	Total	15	4	7	4
	Jumlah Barang				
	Tinggi	3	0	0	3
	Sedang	10	2	7	1
	Sedikit	2	2	0	0
	Kegiatan Promosi				
	Brosur Produk (BP)	3	2	1	0
	Promosi Media social (PMS)	5	1	3	1
	Pemberian diskon (PD)	4	0	3	1
	Toko (TA)	3	1	0	2
	Kualitas				
	Tinggi	6	1	2	3
	Sedang	5	0	4	1
	Rendah	4	2	2	0

a. Nilai Entropy(S)

$$\text{Rumus Entropy}(S) = \sum_{i=1}^n - p_i * \log_2 p_i$$

$$\text{Total} = \left(-\frac{4}{15} * \log_2 \frac{4}{15}\right) + \left(-\frac{7}{15} * \log_2 \frac{7}{15}\right) + \left(-\frac{4}{15} * \log_2 \frac{4}{15}\right) = 1,5301$$

$$\text{Jumlah Barang}_{(\text{Banyak})} = \left(-\frac{0}{3} * \log_2 \frac{0}{3}\right) + \left(-\frac{0}{3} * \log_2 \frac{0}{3}\right) + \left(-\frac{3}{3} * \log_2 \frac{3}{3}\right) = 0$$

$$\text{Jumlah Barang}_{(\text{Sedang})} = \left(-\frac{2}{10} * \log_2 \frac{2}{10}\right) + \left(-\frac{7}{10} * \log_2 \frac{7}{10}\right) + \left(-\frac{1}{10} * \log_2 \frac{1}{10}\right) = 1,1567$$

$$\text{Jumlah Barang}_{(\text{Sedikit})} = \left(-\frac{2}{2} * \log_2 \frac{2}{2}\right) + \left(-\frac{0}{2} * \log_2 \frac{0}{2}\right) + \left(-\frac{0}{2} * \log_2 \frac{0}{2}\right) = 0$$

$$\text{Kegiatan Promosi}_{(\text{BP})} = \left(-\frac{2}{3} * \log_2 \frac{2}{3}\right) + \left(-\frac{1}{3} * \log_2 \frac{1}{3}\right) + \left(-\frac{0}{3} * \log_2 \frac{0}{3}\right) = 0$$

$$\text{Kegiatan Promosi}_{(\text{PMS})} = \left(-\frac{1}{5} * \log_2 \frac{1}{5}\right) + \left(-\frac{3}{5} * \log_2 \frac{3}{5}\right) + \left(-\frac{1}{5} * \log_2 \frac{1}{5}\right) = 0,4421$$

$$\text{Kegiatan Promosi}_{(\text{PD})} = \left(-\frac{0}{4} * \log_2 \frac{0}{4}\right) + \left(-\frac{3}{4} * \log_2 \frac{3}{4}\right) + \left(-\frac{1}{4} * \log_2 \frac{1}{4}\right) = 0$$

$$\text{Kegiatan Promosi}_{(\text{TK})} = \left(-\frac{1}{3} * \log_2 \frac{1}{3}\right) + \left(-\frac{0}{3} * \log_2 \frac{0}{3}\right) + \left(-\frac{2}{3} * \log_2 \frac{2}{3}\right) = 0$$

$$\text{Kualitas}_{(\text{Tinggi})} = \left(-\frac{1}{6} * \log_2 \frac{1}{6}\right) + \left(-\frac{2}{6} * \log_2 \frac{2}{6}\right) + \left(-\frac{3}{6} * \log_2 \frac{3}{6}\right) = 1,4591$$

$$\text{Kualitas}_{(\text{Sedang})} = \left(-\frac{0}{5} * \log_2 \frac{0}{5}\right) + \left(-\frac{4}{5} * \log_2 \frac{4}{5}\right) + \left(-\frac{1}{5} * \log_2 \frac{1}{5}\right) = 0$$

$$\text{Kualitas}_{(\text{Rendah})} = \left(-\frac{2}{4} * \log_2 \frac{2}{4}\right) + \left(-\frac{2}{4} * \log_2 \frac{2}{4}\right) + \left(-\frac{0}{4} * \log_2 \frac{0}{4}\right) = 0$$

b. Gain

$$\text{Rumus Gain}(S,A) = \text{Entropy}(S) - \sum_{i=1}^n - \frac{|S_i|}{|S|} * \text{Entropy}(S_i)$$

$$\text{Jumlah Barang} = 1,5301 - \left(\left(\frac{3}{15} * 0 \right) + \left(\frac{10}{15} * 1,1567 \right) + \left(\frac{2}{15} * 0 \right) \right) = 0,7589$$

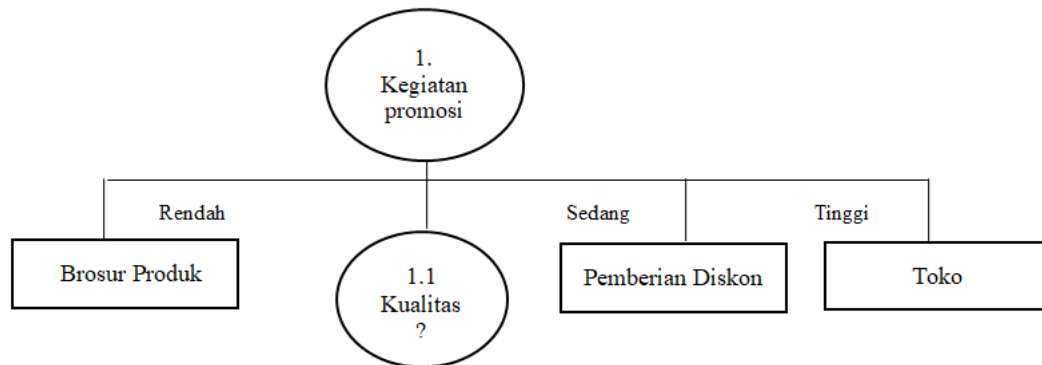
$$\text{Kegiatan Promosi} = 1,5301 - \left(\left(\frac{3}{15} * 0 \right) + \left(\frac{5}{15} * 0,4421 \right) + \left(\frac{4}{15} * 0 \right) + \left(\frac{3}{15} * 0 \right) \right) = 1,3827$$

$$\text{Kualitas} = 1,5301 - \left(\left(\frac{6}{15} * 1,4591 \right) + \left(\frac{5}{15} * 0 \right) + \left(\frac{4}{15} * 0 \right) \right) = 0,9464$$

Tabel 5. Rekapitulasi Hasil Node awal

Node	Keterangan	Jumlah Kasus	Rendah	Sedang	Tinggi	Entropy	Gain
1	Total	15	4	7	4	1,5301	
	Jumlah Barang						0,7589
	Tinggi	3	0	0	3	0	
	Sedang	10	2	7	1	1,1567	
	Sedikit	2	2	0	0	0	
	Kegiatan Promosi						1,3827
	Brosur Produk (BP)	3	2	1	0	0	
	Promosi Media social (PMS)	5	1	3	1	0,4421	
	Pemberian discon (PD)	4	0	3	1	0	
	Toko (Tk)	3	1	0	2	0	
	Kualitas						0,9464
	Tinggi	6	1	2	3	1,4591	
	Sedang	5	0	4	1	0	
	Rendah	4	2	2	0	0	

Kedua: Pembentukan akar pohon keputusan berdasarkan hasil dari Gain tertinggi. Berdasarkan perhitungan Gain di atas diperoleh dari Gain tertinggi yaitu dengan attribute Kegiatan promosi, Sehingga Kebutuhan menjadi node akar awal pembentukan pohon keputusan.



Gambar 4. Pohon Keputusan Node Awal

Setelah diketahui data yang tidak diketahui nilainya, selanjutnya dilakukan pembentukan data baru dimana hanya data “Promosi Media Sosial” yang dilakukan perhitungan kembali, adapun data baru dapat dilihat sebagai berikut ini:

Tabel 6. Sampel Data Setelah Node Awal

No	Bulan	Jumlah Barang	Kegiatan Promosi	Kualitas	Pola Penjualan
1	Maret	Sedang	Promosi sosial media	Tinggi	Sedikit
2	Juni	Sedang	Promosi sosial media	Sedang	Sedang
3	November	Sedang	Promosi sosial media	Sedang	Sedang
4	Desember	Sedang	Promosi sosial media	Rendah	Sedang
5	Februari	Sedang	Promosi sosial media	Sedang	Tinggi

Lakukan perhitungan awal kembali, sehingga data berikut disebut dengan node 1.1 adapun hasil dari perhitungan sebagai berikut ini:

Tabel 7. Rekapitulasi Hasil Node awal

Node	Keterangan	Jumlah Kasus	Rendah	Sedang	Tinggi	Entropy	Gain
1	Kegiatan Promosi = Promosi sosial media	5	1	3	1	1,3709	
	Jumlah Barang						0
	Banyak	0	0	0	0	0	
	Sedang	5	1	3	1	1,3709	

Node	Keterangan	Jumlah Kasus	Rendah	Sedang	Tinggi	Entropy	Gain
	Sedikit	0	0	0	0	0	
	Kualitas						1,3709
	Tinggi	1	1	0	0	0	
	Sedang	3	0	2	1	0	
	Rendah	1	0	1	0	0	

c. Nilai Entropy(S)

$$\text{Rumus Entropy}(S) = \sum_{i=1}^n - p_i * \log_2 p_i$$

$$\text{Kegiatan Promosi}_{(PMS)} = \left(-\frac{1}{5} * \log_2 \frac{1}{5}\right) + \left(-\frac{3}{5} * \log_2 \frac{3}{5}\right) + \left(-\frac{1}{5} * \log_2 \frac{1}{5}\right) = 1,3709$$

$$\text{Jumlah Barang}_{(Banyak)} = \left(-\frac{0}{0} * \log_2 \frac{0}{0}\right) + \left(-\frac{0}{0} * \log_2 \frac{0}{0}\right) + \left(-\frac{0}{0} * \log_2 \frac{0}{0}\right) = 0$$

$$\text{Jumlah Barang}_{(Sedang)} = \left(-\frac{1}{5} * \log_2 \frac{1}{5}\right) + \left(-\frac{3}{5} * \log_2 \frac{3}{5}\right) + \left(-\frac{1}{5} * \log_2 \frac{1}{5}\right) = 1,3709$$

$$\text{Jumlah Barang}_{(Sedikit)} = \left(-\frac{0}{0} * \log_2 \frac{0}{0}\right) + \left(-\frac{0}{0} * \log_2 \frac{0}{0}\right) + \left(-\frac{0}{0} * \log_2 \frac{0}{0}\right) = 0$$

$$\text{Kualitas}_{(Tinggi)} = \left(-\frac{1}{1} * \log_2 \frac{1}{1}\right) + \left(-\frac{0}{1} * \log_2 \frac{0}{1}\right) + \left(-\frac{0}{1} * \log_2 \frac{0}{1}\right) = 0$$

$$\text{Kualitas}_{(Sedang)} = \left(-\frac{0}{3} * \log_2 \frac{0}{3}\right) + \left(-\frac{2}{3} * \log_2 \frac{2}{3}\right) + \left(-\frac{1}{3} * \log_2 \frac{1}{3}\right) = 0$$

$$\text{Kualitas}_{(Rendah)} = \left(-\frac{1}{1} * \log_2 \frac{1}{1}\right) + \left(-\frac{1}{1} * \log_2 \frac{1}{1}\right) + \left(-\frac{0}{1} * \log_2 \frac{0}{1}\right) = 0$$

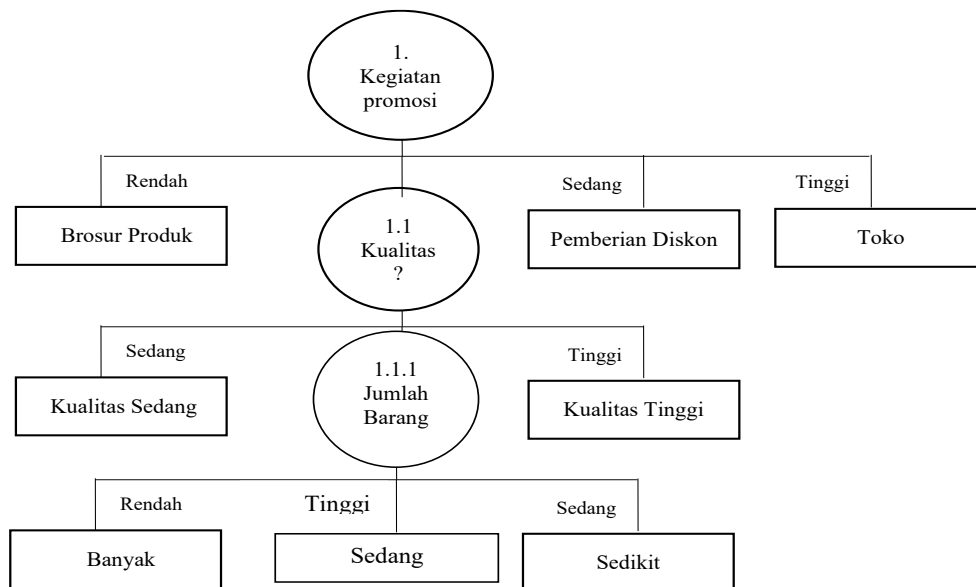
d. Gain

$$\text{Rumus Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n - \frac{|S_i|}{|S|} * \text{Entropy}(S_i)$$

$$\text{Jumlah Barang} = 1,3709 - \left(\left(\frac{0}{5} * 0\right) + \left(\frac{5}{5} * 1,3709\right) + \left(\frac{0}{5} * 0\right)\right) = 0$$

$$\text{Kualitas} = 1,3709 - \left(\left(\frac{1}{5} * 0\right) + \left(\frac{3}{5} * 0\right) + \left(\frac{1}{5} * 0\right)\right) = 1,3709$$

Dari hasil di atas maka terbentuk pohon keputusan Node 1.1 hingga 1.1.1 dikarenakan sudah tidak ada nilai yang harus dihitung sehingga pohon keputusan terbentuk sebagai berikut ini:



Gambar 5. Pohon Keputusan Akhir

Pohon keputusan yang telah dibentuk dibaca dalam bentuk rule/aturan. Cara membaca pohon keputusan dimulai dari node akar ke node cabang menggunakan kata “DAN” untuk penghubung antara node dan dibentuk dalam Jika...maka... atau Jika...dan...maka... Bentuk rule pohon keputusan di atas adalah sebagai berikut:

1. Jika Kegiatan Promosi = Brosur Produk maka Hasil = Rendah
2. Jika Kegiatan Promosi = Pemberian Diskon maka hasil = Sedang
3. Jika Kegiatan Promosi = Tidak Ada maka hasil = Tinggi
4. Jika Kegiatan Promosi = Pemberian Diskon dan Kualitas = Sedang maka hasil = Rendah
5. Jika Kegiatan Promosi = Pemberian Diskon dan Kualitas = Tinggi maka hasil = Sedang
6. Jika Kegiatan Promosi = Pemberian Diskon, Kualitas dan Jumlah Barang = Banyak maka hasil = Rendah
7. Jika Kegiatan Promosi = Pemberian Diskon, Kualitas dan Jumlah Barang = Sedang maka hasil = Tinggi
8. Jika Kegiatan Promosi = Pemberian Diskon, Kualitas dan Jumlah Barang = Sedikit maka hasil = Sedang

Dari hasil yang ditemukan maka dapat diambil kesimpulan bahwa pola penjualan springbed dapat terjadi tiga kondisi sebagai berikut ini: Penjualan springbed dapat diklasifikasikan menjadi tiga kategori berdasarkan tingkat pendapatan dan faktor yang memengaruhinya, yaitu penjualan rendah, sedang, dan tinggi. Penjualan rendah terjadi ketika pendapatan berada di kisaran sekitar 2 miliar rupiah, yang umumnya dipengaruhi oleh strategi promosi yang kurang efektif seperti hanya mengandalkan penyebaran brosur dan pemberian diskon, serta didukung oleh kualitas produk yang sedang dan ketersediaan stok yang terbatas, sehingga minat konsumen cenderung rendah dan berdampak pada penurunan penjualan. Sementara itu, penjualan sedang berada pada rentang di atas 2,5 miliar hingga 5 miliar rupiah, yang biasanya disebabkan oleh adanya keseimbangan antara pemberian diskon dan jumlah stok yang tersedia, sehingga penjualan tetap berjalan namun tidak menunjukkan peningkatan yang signifikan. Adapun penjualan tinggi terjadi pada kisaran di atas 5 miliar hingga 10 miliar rupiah, yang didorong oleh strategi promosi yang lebih luas dan efektif, seperti distribusi produk ke berbagai toko, didukung oleh kualitas produk yang unggul dan inovatif, serta ketersediaan stok yang memadai, sehingga mampu menarik minat konsumen dalam jumlah besar, termasuk dari luar daerah, dan menghasilkan peningkatan permintaan yang signifikan. **Pengujian Hasil**

Setelah dilakukan proses perhitungan manual dengan menggunakan algoritma C4.5 Decision Tree, maka selanjutnya dilakukan proses pengujian. Dimana, pengujian terhadap hasil dan algoritma dengan menggunakan Tools Rapid Miner Studio. Adapun hasil dapat dilihat:

accuracy: 60.00%

	true Tinggi	true Sedikit	true Sedang	class precision
pred. Tinggi	2	1	0	66.67%
pred. Sedikit	0	2	1	66.67%
pred. Sedang	4	0	5	55.56%
class recall	33.33%	66.67%	83.33%	

Gambar 6. Hasil Pengujian

Hasil pengujian model klasifikasi menggunakan algoritma Decision Tree menunjukkan tingkat akurasi sebesar 60%, yang mengindikasikan bahwa model masih memiliki kemampuan prediksi yang sedang dalam mengklasifikasikan data penjualan ke dalam tiga kelas, yaitu Tinggi, Sedang, dan Sedikit. Berdasarkan confusion matrix, pada kelas aktual Tinggi, model berhasil mengklasifikasikan dengan benar sebanyak 2 data, namun masih terdapat 4 data yang salah diprediksi menjadi kelas Sedang dan 1 data menjadi Sedikit, sehingga menghasilkan nilai recall sebesar 33,33%. Pada kelas Sedikit, model mampu mengklasifikasikan 2 data dengan benar, dengan 1 data salah diprediksi sebagai Tinggi dan 1 data sebagai Sedang, sehingga recall mencapai 66,67%. Sementara itu, pada kelas Sedang, model menunjukkan performa terbaik dengan 5 data terklasifikasi benar, 1 data salah ke Sedikit, dan tidak ada yang salah ke Tinggi, sehingga menghasilkan recall tertinggi yaitu 83,33%.

Dari sisi precision, kelas Tinggi dan Sedikit masing-masing memperoleh nilai 66,67%, sedangkan kelas Sedang memiliki precision sebesar 55,56%. Hal ini menunjukkan bahwa meskipun model cukup baik dalam mengenali beberapa pola data, masih terdapat ketidakseimbangan dalam distribusi prediksi antar kelas, khususnya pada kelas Tinggi yang memiliki tingkat recall paling rendah. Kondisi ini mengindikasikan bahwa model belum sepenuhnya optimal dan masih dapat ditingkatkan, misalnya melalui penambahan data latih, balancing dataset, atau tuning parameter pada algoritma C4.5 agar hasil klasifikasi menjadi lebih akurat dan stabil.

4. KESIMPULAN

Berdasarkan hasil dari memprediksi omset penjualan springbed pada menggunakan Algoritma C4.5, maka dapat disimpulkan bahwa proses dalam pencarian penjualan tertinggi pada penjualan springbed dengan tidak dilakukannya promosi seperti pemberian diskon yang membuat harga menjadi rendah, memberikan produk berkualitas tinggi sehingga pendapatan di atas 8 miliar. Penerapan C4.5 membentuk pohon keputusan yang mudah dipahami berdasarkan rule kasus yang ada sehingga membangun strategi baru dari hasil penelitian untuk meningkatkan harga penjualan. Penerapan Algoritma C4.5 dalam mempercepat hasil prediksi dengan akurasi sebesar 60%

REFERENCES

- [1] F. Akbar, H. W. Saputra, A. K. Maulaya, M. F. Hidayat, and R. Rahmaddeni, "Implementasi Algoritma Decision Tree C4.5 dan Support Vector Regression untuk Prediksi Penyakit Stroke," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 2, pp. 61–67, 2022, doi: <https://doi.org/10.57152/malcom.v2i2.426>.
- [2] U. Suriani, "Penerapan Data Mining untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4.5," *J. Comput. Inf. Syst. Ampera*, vol. 4, no. 2, pp. 55–65, 2023, doi: <https://doi.org/10.51519/journalcisa.v4i2.393>.
- [3] A. Rahim and P. A. Jusia, "Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Algoritma Decision Tree Classifier," *Indones. J. Comput. Sci.*, vol. 13, no. 1, 2024, doi: <https://doi.org/10.33022/ijcs.v13i1.3692>.

- [4] A. Rahmat and I. Permana, "Prediksi Waktu Panen Jamur Tiram Berdasarkan Data Lingkungan Kubung Menggunakan Metode Data Mining Algoritma Decision Tree," *RIGGS J. Artif. Intell. Digit. Bus.*, vol. 5, no. 1, pp. 2817–2828, 2026, doi: <https://doi.org/10.31004/riggs.v5i1.6418>.
- [5] E. Sabna and O. Dewi, "Prediksi Penyakit Stroke menggunakan Algoritma Decision Tree dan Naïve Bayes," *RIGGS J. Artif. Intell. Digit. Bus.*, vol. 4, no. 3, pp. 1294–1299, 2025, doi: <https://doi.org/10.31004/riggs.v4i3.2132>.
- [6] M. F. Aditya, A. Pramuntadi, D. P. Wijaya, and Y. Wicaksono, "Implementasi Metode Decision Tree pada Prediksi Penyakit Diabetes Melitus Tipe 2," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 3, pp. 1104–1110, 2024, doi: <https://doi.org/10.57152/malcom.v4i3.1284>.
- [7] I. Syafii, A. A. Ribhi, L. Y. Astutik, G. K. S. Budiono, and A. S. Pamela, "Analisis Prediksi Customer Repeat Order menggunakan Algoritma Decision Tree pada Perusahaan Transportasi," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 4, pp. 1372–1378, 2024, doi: <https://doi.org/10.57152/malcom.v4i4.1538>.
- [8] A. K. Bhatnagar, P. Madan, A. Rana, S. Sharma, S. Sonawane, and C. Vimala Josphine, "An Efficient Techniques For Disease Prediction From Medical Data Using Data Mining And Machine Learning," in *International Conference on Contemporary Computing and Informatics (IC3I)*, 2025. doi: 10.1109/IC3I56241.2022.10073286.
- [9] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," *Smart Learn. Environ.*, vol. 9, no. 11, 2022, doi: <https://doi.org/10.1186/s40561-022-00192-z>.
- [10] Y. M. Dawa, A. Aziz, and M. Ahsan, "OPTIMASI ALGORITMA C4.5 BERBASIS PARTICLE SWARM OPTIMIZATION (PSO) UNTUK MENENTUKAN WHOLESALERS PENJUALAN," *BIMASAKTI*, vol. 6, no. 1, pp. 21–26, 2023, doi: 10.21067/bimasakti.v6i1.9216.
- [11] A. U. Rehman, R. M. Saleem, Z. Shafi, M. Imran, M. Pradhan, and H. M. Alzoubi, "Analysis of Income on the Basis of Occupation using Data Mining," in *International Conference on Business Analytics for Technology and Security (ICBATS)*, 2022. doi: 10.1109/ICBATS54253.2022.9759040.
- [12] V. S. K. Reddy, P. Meghana, N. V. S. Reddy, and B. A. Rao, "Prediction on Cardiovascular disease using Decision tree and Naïve Bayes classifiers," *J. Phys. Conf. Ser.*, vol. 2161, 2026, doi: 10.1088/1742-6596/2161/1/012015.
- [13] S. A. Narendra, P. A. Setiawan, S. Anwar, and M. H. Fuad, "Analisa Penjualan Rokok Dengan Metode Klasifikasi Menggunakan Formula Algoritma C4.5 Pada CV Jaya Berkah Mas," *J. Insa. J. Inf. Syst. Manag. Innov.*, vol. 4, no. 2, 2024, doi: <https://doi.org/10.31294/jinsan.v4i2.4656>.
- [14] A. A. Dermawan, K. Vadilla, and A. M. Leman, "Optimasi Klasifikasi Penjualan Produk Skincare Terlaris di Klinik XYZ Menggunakan Algoritma Decision Tree C4.5," *J. Pendidik. Sains dan Komput.*, vol. 5, no. 1, p. <https://doi.org/10.47709/jpsk.v5i01.5529>, 2025.
- [15] B. Triawan, I. Lubis, and L. A. N. Kadim, "Penerapan Data Mining Untuk Prediksi Penjualan Spanduk Menggunakan Algoritma C4.5," *J. MATECH*, vol. 3, no. 2, 2023, doi: <https://doi.org/10.63893/matech.v3i2.172>.
- [16] Rismayanti, M. N. Aidi, and H. Wijayanto, "Penerapan Algoritma C4.5 dan Random Forest pada Tingkat Penjualan Serum Somethinc di Shopee," *Xplore J. Stat.*, vol. 12, no. 3, pp. 263–273, 2023, doi: <https://doi.org/10.29244/xplore.v12i3.1150>.
- [17] M. Saberi-Karimian *et al.*, "Data mining approaches for type 2 diabetes mellitus prediction using anthropometric measurements," *J. Clin. Lab. Anal.*, vol. 37, no. 1, 2022, doi: <https://doi.org/10.1002/jcla.24798>.
- [18] S. Batool, J. Rashid, M. W. Nisar, J. Kim, H.-Y. Kwon, and A. Hussain, "Educational data mining to predict students' academic performance: A survey study," *Educ. Inf. Technol.*, vol. 28, pp. 905–971, 2023, doi: <https://doi.org/10.1007/s10639-022-11152-y>.
- [19] A. Simanjorang, A. M. H. Pardede, and S. Syahputra, "Implementasi Algoritma C4.5 untuk memprediksi Penjualan Paket Internet," *KAKIFIKOM*, vol. 6, no. 2, pp. 106–115, 2024, doi: <https://doi.org/10.54367/kakifikom.v6i2.4093>.
- [20] M. J. Ruliansyah and M. Betty, "Penerapan Metode C4.5 dalam Prediksi Penjualan Tim Bev 1 pada PT. Surya Pangan Sejahtera Bekasi Jawa Barat," *JIMIK*, vol. 5, no. 2, pp. 1269–1278, 2024, doi: 10.35870/jmik.v5i2.664.
- [21] K. Arumugam, M. Naved, P. P. Shinde, O. Leiva-Chauca, A. Huaman-Osorio, and T. Gonzales-Yanac, "Multiple disease prediction using Machine learning algorithms," *Mater. Today Proc.*, vol. 80, no. 3, pp. 3682–3685, 2023, doi: <https://doi.org/10.1016/j.matpr.2021.07.361>.
- [22] Calvin and R. Harman, "PENERAPAN ALGORITMA C4.5 UNTUK MEMPREDIKSI PROFIT PENJUALAN," *COMASIE*, vol. 7, no. 2, pp. 52–62, 2022, doi: 10.33884/comasiejournal.v7i2.5974.
- [23] K. Eldora, E. Fernando, and W. Winanti, "Comparative Analysis of KNN and Decision Tree Classification Algorithms for Early Stroke Prediction: A Machine Learning Approach," *J. Inf. Syst. Informatics*, vol. 6, no. 1, pp. 313–338, 2024, doi: <https://doi.org/10.51519/journalisi.v6i1.664>.
- [24] H. Pallathadka, A. Wenda, E. Ramirez-Asis, M. Asís-López, J. Flores-Albornoz, and K. Phasinam, "Classification and prediction of student performance data using various machine learning algorithms," *Mater. Today Proc.*, vol. 80, no. 3, pp. 3782–3785, 2023, doi: <https://doi.org/10.1016/j.matpr.2021.07.382>.
- [25] M. T. Huyut and H. Üstündağ, "Prediction of diagnosis and prognosis of COVID-19 disease by blood gas parameters using decision trees machine learning model a retrospective observational study," *Med. Gas Res.*, vol. 12, no. 2, pp. 60–66, 2022, doi: 10.4103/2045-9912.326002.
- [26] H. Blockeel, L. Devos, B. Frénay, G. Nanfack, and S. Nijssen, "Decision trees: from efficient prediction to responsible AI," *Mach. Learn. Artif. Intell.*, vol. 6, 2023, doi: <https://doi.org/10.3389/frai.2023.1124553>.
- [27] Y. F. Saputra and Sugiarto, "Algoritma C4.5 sebagai Penerapan Decision Tree-Based Classification Model untuk Mengklasifikasikan Tingkat Omzet UMKM Berdasarkan Profil Bisnis," *J. Ekon. Manaj. Akunt.*, vol. 13, no. 2, pp. 188–197, 2024, doi: 10.24903/je.v13i2.3562.
- [28] 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," *Int. J. Softw. Eng. Comput. Sci.*, vol. 4, no. 3, pp. 943–952, 2024, doi: <https://doi.org/10.35870/ijsecs.v4i3.2999>.
- [29] F. G. Falentina, G. Y. Y. Wabdaron, D. Andiyani, M. S. Wondiwoi, and H. Sutejo, "Analysis of Data Mining Implementation for Classifying Best-Selling Food Sale Using the Decision Tree (C4.5) Algorithm," *J. Ilm. Sist. Inf.*, vol. 4, no. 2, pp. 88–97, 2025, doi: <https://doi.org/10.51903/f2jegk76>.



- [30] J. R. Sagala, P. S. Hasugian, M. Van Breukelen, and W. S. U. Nantheni, "ANALYSIS OF THE C4.5 ALGORITHM IN PREDICTING FERTILIZER AND PESTICIDE SALES AT UD. MR ZEN AGRO," *J. Sist. Inf. dan Ilmu Komput.*, vol. 7, no. 2, pp. 26–41, 2024, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v7i2.4816.