

Penerapan Algoritma C4.5 dalam Klasifikasi Tingkat Konsumsi Harian Masyarakat untuk Mengurangi Food Waste pada UMKM di Kota Medan

Dea Khairani*, Puji Sari Ramadhan, Rina Mahyuni

Sistem Informasi, STMIK TRIGUNA DHARMA, Medan, Indonesia

Email: ¹deakhairani544@gmail.com, ²pujisariramadhan@gmail.com, ³rinamahyuni14@gmail.com

Email Penulis Korespondensi: deakhairani544@gmail.com

Submitted 19-03-2026; Accepted 06-04-2026; Published 28-04-2026

Abstrak

Penelitian ini membahas permasalahan ketidaktepatan penentuan jumlah produksi pada UMKM sarapan pagi di Kota Medan akibat fluktuasi permintaan, yang sering menimbulkan sisa makanan (food waste) dan kerugian. Tujuan penelitian adalah membangun model klasifikasi tingkat konsumsi harian (tinggi, sedang, rendah) sebagai dasar rekomendasi produksi. Metode yang digunakan adalah algoritma Decision Tree C4.5 dengan dataset sebanyak 92 catatan operasional harian yang mencakup atribut kategori produksi, jenis menu, kondisi cuaca, dan hari operasional. Data di proses melalui pengkategorian atribut, kemudian dianalisis menggunakan perhitungan entropy dan gain ratio untuk membentuk pohon keputusan. Evaluasi model dilakukan menggunakan 10-fold cross validation pada RapidMiner. Hasil penelitian menunjukkan bahwa atribut kategori produksi memiliki nilai gain ratio tertinggi sehingga menjadi akar pohon keputusan. Model yang dihasilkan mencapai akurasi sebesar 57,56% dengan deviasi $\pm 7,85\%$, dengan performa terbaik pada kelas konsumsi sedang. Kontribusi utama penelitian ini adalah menghasilkan aturan keputusan berbasis IF-THEN yang dapat digunakan secara praktis oleh UMKM untuk menyesuaikan jumlah produksi harian berdasarkan kondisi operasional, sehingga membantu mengurangi potensi food waste tanpa memerlukan perhitungan yang kompleks.

Kata Kunci: C4.5; Tingkat Konsumsi; Rekomendasi Produksi; Food Waste; UMKM

Abstract

This study addresses the issue of inaccurate production forecasting among breakfast-serving SMEs in Medan due to fluctuations in demand, which often result in food waste and financial losses. The objective of this study is to develop a classification model for daily consumption levels (high, medium, low) to serve as the basis for production recommendations. The method used is the C4.5 Decision Tree algorithm with a dataset of 92 daily operational records covering the attributes of production category, menu type, weather conditions, and operational days. The data was preprocessed through attribute categorization, then analyzed using entropy and gain ratio calculations to form a decision tree. Model evaluation was performed using 10-fold cross-validation in RapidMiner. The results showed that the production category attribute had the highest gain ratio, making it the root of the decision tree. The resulting model achieved an accuracy of 57.56% with a deviation of $\pm 7.85\%$, performing best in the moderate consumption class. The primary contribution of this study is the generation of IF-THEN-based decision rules that can be practically applied by SMEs to adjust daily production volumes based on operational conditions, thereby helping to reduce potential food waste without requiring complex calculations.

Keywords: C4.5; Consumption Level; Production Recommendation; Food Waste; MSMEs

1. PENDAHULUAN

UMKM kuliner, khususnya usaha sarapan pagi, menjadi komponen vital dalam lanskap bisnis lokal di Kota Medan karena peranannya dalam memenuhi kebutuhan makanan harian masyarakat perkotaan. Namun kenyataannya, banyak pelaku usaha kecil ini masih menghadapi kesulitan besar dalam memperkirakan jumlah produksi harian secara akurat. Fluktuasi permintaan membuat mereka kerap menyiapkan terlalu banyak makanan yang tidak terjual, atau sebaliknya, kekurangan stok saat permintaan melonjak. Ketidaksesuaian antara jumlah produksi dan tingkat konsumsi tersebut secara langsung berkontribusi terhadap meningkatnya sisa makanan (food waste) pada usaha sarapan pagi. Kondisi tersebut menimbulkan pemborosan makanan yang tidak sedikit, sekaligus kerugian finansial bagi pemilik UMKM.

Permasalahan ini tidak hanya bersifat operasional tetapi juga strategis. Mengandalkan intuisi dan pengalaman semata dalam pengambilan keputusan produksi dapat mengakibatkan inefisiensi yang sistemik. Untuk memberikan solusi terhadap isu ini, penelitian dengan pendekatan *data mining* layak diterapkan. Di bidang teknologi informasi, pemanfaatan data mining semakin penting karena mampu memberikan gambaran yang lebih mendalam mengenai perilaku dan pola bisnis. Pendekatan ini sejalan dengan pernyataan bahwa data mining berfungsi untuk menemukan pola tersembunyi, mengekstraksi pengetahuan, dan membantu proses pengambilan keputusan berbasis klasifikasi[1]. Salah satu algoritma klasifikasi yang banyak digunakan adalah C4.5. Algoritma ini membangun pohon keputusan dengan memilih atribut yang memiliki nilai information gain tertinggi sebagai dasar pemisahan data. Struktur pohon yang dihasilkan memungkinkan pembentukan aturan keputusan yang sistematis dan transparan. Dalam konteks penelitian ini, kemampuan tersebut penting karena klasifikasi tingkat konsumsi harian tidak hanya berfungsi sebagai

hasil prediksi, tetapi juga sebagai dasar rekomendasi produksi untuk mendukung pengendalian food waste pada UMKM sarapan pagi.

Berbagai penelitian baru di ranah UMKM, produksi kuliner, dan sampah makanan telah membahas isu-isu terkait. Sebagai contoh, bahwa UMKM kuliner di Medan menghadapi tekanan kuat dari waralaba asing yang memperkuat perlunya strategi adaptif lokal[2]. Penelitian terdahulu menunjukkan bahwa UMKM tidak hanya perlu berinovasi secara produk, tetapi juga meningkatkan kemampuan manajerial agar dapat bersaing[3]. Sementara itu, dalam penelitian lainnya menyebutkan tentang potensi gerakan anti-food waste sebagai salah satu cara memulihkan daya ekonomi UMKM, dengan memanfaatkan sisa makanan sebagai kesempatan distribusi ulang[4]. Hal ini mendukung urgensi analisis pola konsumsi agar produksi tidak berlebih dan sisa makanan bisa diminimalisir.

Dari sisi dampak ekonomi nasional, potensi kerugian akibat food loss dan food waste di Indonesia pun sangat besar. Hasil kajian empiris di Indonesia menunjukkan bahwa tingginya angka *food loss* dan *food waste* tidak hanya menimbulkan dampak negatif terhadap lingkungan, tetapi juga menyebabkan kerugian ekonomi yang cukup besar. Beberapa studi memperkirakan bahwa nilai kerugian akibat pemborosan pangan tersebut dapat mencapai sekaitar Rp 107–346 triliun per tahun, sehingga pengelolaan sampah makanan secara lebih efektif menjadi hal yang sangat penting dalam mendukung efisiensi sistem pangan nasional[5]. Hal ini menegaskan bahwa tidak hanya UMKM yang dirugikan, tetapi perekonomian lebih luas juga terpengaruh. Penelitian lain di level teoretis, yang memperlihatkan bahwa praktik food waste di UMKM dapat dikaji dari perspektif material flow cost accounting (MFCA) dan teori netralisasi perilaku agar usaha kecil dapat mengidentifikasi aliran sampah makanan dan membuat mitigasi yang lebih efektif [6].

Di ranah teknologi, data mining telah terbukti bermanfaat dalam memberikan wawasan bisnis yang mendalam. Penelitian menunjukkan bahwa algoritma C4.5 dapat digunakan untuk mengklasifikasikan strategi penjualan produk UMKM dengan akurasi tinggi, sehingga membantu UMKM dalam merumuskan strategi penjualan yang lebih tepat. Selain itu, penelitian lainnya menunjukkan pengaruh positif dari analitik data besar (*big data analytics*) berbasis C4.5 terhadap keputusan strategis dan kinerja pemasaran UMKM di era digital [7].

Meskipun penelitian-penelitian di atas sangat relevan, masih terdapat celah yang cukup signifikan. Pertama, belum ada penelitian yang secara spesifik mengkombinasikan analisis data historis konsumsi makanan harian dengan klasifikasi tingkat konsumsi (tinggi, sedang, rendah) di segmen UMKM sarapan pagi di kota seperti Medan. Kedua, meskipun ada penelitian terkait food waste dan upaya manajemen sampah, sedikit yang mengaitkannya langsung dengan keputusan produksi harian berbasis model klasifikasi. Ketiga, dalam banyak studi data mining pada UMKM, fokusnya lebih ke prediksi penjualan atau stok jangka panjang; belum banyak yang mengeksplorasi prediksi konsumsi harian untuk menentukan produksi saat itu juga.

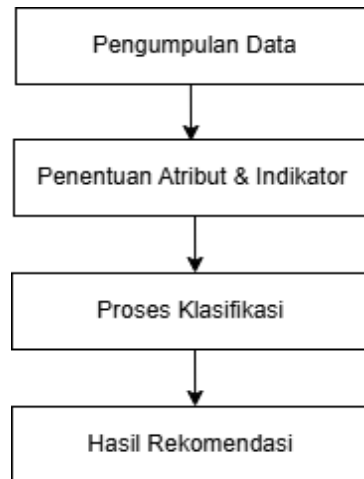
Di sisi lain, meskipun terdapat berbagai algoritma klasifikasi seperti Naïve Bayes dan K-Nearest Neighbor (KNN) yang dikenal memiliki keunggulan dalam efisiensi perhitungan dan kemampuan mengenali pola data, pendekatan tersebut umumnya kurang menekankan aspek interpretabilitas yang penting bagi pelaku UMKM. Oleh karena itu, penelitian ini mengusulkan penggunaan algoritma Decision Tree C4.5 yang tidak hanya mampu melakukan klasifikasi, tetapi juga menghasilkan aturan keputusan yang mudah dipahami. Dengan demikian, kontribusi utama (*novelty*) penelitian ini terletak pada integrasi model klasifikasi tingkat konsumsi harian dengan rekomendasi produksi berbasis aturan IF THEN yang aplikatif, sehingga dapat digunakan secara langsung oleh pelaku UMKM untuk mengurangi food waste, meningkatkan efisiensi produksi, serta mendukung pengambilan keputusan harian tanpa memerlukan proses analisis yang kompleks.

Berdasarkan analisis gap tersebut, penelitian ini bertujuan membangun model klasifikasi tingkat konsumsi harian menggunakan algoritma C4.5 serta menghasilkan aturan keputusan sebagai dasar rekomendasi produksi untuk mendukung pengendalian food waste pada UMKM sarapan pagi di Kota Medan. Model ini akan menggunakan atribut operasional seperti jumlah produksi, penjualan harian, jenis menu, hari operasional, dan data historis lainnya sebagai input. Harapannya, dengan model ini, pemilik UMKM dapat membuat keputusan produksi yang lebih akurat, mengurangi sisa makanan, dan sekaligus meningkatkan efisiensi usaha serta profitabilitas. Selain itu, hasil penelitian juga diharapkan dapat menjadi dasar pembuatan aplikasi atau sistem pendukung keputusan sederhana yang mudah diadopsi oleh UMKM kecil tanpa kebutuhan sumber daya analitik yang tinggi.

2. METODOLOGI PENELITIAN

2.1 Kerangka Kerja

Penelitian ini mengikuti alur kerja yang disusun secara sistematis untuk memastikan proses analisis berjalan terarah dan menghasilkan model klasifikasi yang akurat. Kerangka kerja tersebut terdiri dari empat tahapan utama, yaitu :



Gambar 1. Kerangka Kerja

Gambar 1 merupakan tahapan penelitian yang akan dilakukan pada penelitian ini, di mulai dari pengumpulan data, penentuan atribut dan indikator, proses klasifikasi, dan hasil rekomendasi.

a. Pengumpulan Data

Langkah awal dalam penelitian ini dimulai dengan menghimpun data historis dari salah satu UMKM penyedia sarapan pagi di Kota Medan. Data tersebut mencakup informasi mengenai jumlah produksi, jumlah penjualan, jenis menu, kondisi cuaca hari operasional, serta jumlah makanan yang tersisa setiap hari. Pengumpulan data dilakukan melalui tiga metode: dokumentasi catatan operasional, wawancara singkat dengan pemilik usaha, dan observasi langsung di lokasi. Pendekatan kuantitatif digunakan agar data yang diperoleh mencerminkan kondisi dan aktivitas operasional secara akurat. Penelitian terbaru menunjukkan bahwa akurasi data sangat menentukan kualitas model klasifikasi. Pencatatan operasional yang rapi dan konsisten penting untuk merepresentasikan pola produksi dan penjualan secara nyata. Data operasional yang tepat membantu pelaku UMKM mengambil keputusan lebih efektif serta menyesuaikan strategi dengan cepat terhadap perubahan permintaan pasar[8]. Selain dokumentasi, proses verifikasi data melalui interaksi langsung dengan pelaku usaha juga diperlukan agar informasi yang digunakan dalam analisis benar-benar sesuai dengan kondisi di lapangan[9]. Di sisi lain, observasi lapangan sering kali memberikan gambaran tambahan mengenai perilaku konsumen dan aktivitas operasional yang tidak tercatat dalam dokumen administrasi, namun memiliki pengaruh pada tingkat konsumsi harian [10]. Pentingnya akurasi data ini juga diperkuat oleh temuan lainnya yang menekankan bahwa ketidaktepatan data dapat berdampak langsung pada kesalahan perhitungan produksi dan berpotensi meningkatkan food waste[11].

b. Penentuan Atribut dan Indikator

Pada tahap ini, variabel penelitian ditentukan berdasarkan kondisi operasional yang umum pada UMKM kuliner. Atribut yang digunakan meliputi jumlah produksi, penjualan, jenis menu, cuaca, sisa makanan, serta hari operasional. Setiap atribut kemudian diuraikan menjadi indikator yang lebih rinci agar analisis dengan algoritma C4.5 lebih terarah dan hasil klasifikasinya lebih mudah dipahami. Sejumlah penelitian menunjukkan bahwa ketepatan dalam memilih indikator sangat berpengaruh terhadap kualitas hasil klasifikasi. Misalnya, penelitian lain menemukan bahwa penentuan fitur yang relevan dan disusun secara tepat mampu meningkatkan akurasi algoritma C4.5 ketika digunakan untuk menganalisis data UMKM[2].

Selain itu, studi yang dilakukan oleh beberapa penelitian menunjukkan bahwa pengelolaan food waste berbasis data dapat berjalan jauh lebih efektif ketika indikator-indikator seperti jumlah limbah dan frekuensi kejadiannya dikelompokkan secara jelas. Pengkategorian yang rapi membantu proses analisis menjadi lebih terstruktur dan menghasilkan temuan yang lebih akurat[12].

Sejumlah penelitian lain menunjukkan bahwa algoritma C4.5 mampu mengolah kombinasi atribut numerik dan kategorikal dengan baik. Pada studi klasifikasi jamur, penggunaan karakteristik morfologis berhasil menghasilkan pohon keputusan yang mudah dipahami serta prediksi yang konsisten. Hal ini menunjukkan bahwa data campuran, seperti yang umum pada bidang kuliner, dapat memberikan kinerja model yang kuat ketika dianalisis menggunakan C4.5[13].

c. Proses Klasifikasi

Pada tahap ini, data yang telah dibersihkan dan dikelompokkan dianalisis menggunakan algoritma C4.5. Pembentukan model dilakukan dengan menghitung entropy dan information gain untuk menentukan atribut paling berpengaruh sebagai akar pohon keputusan. Selanjutnya, pohon dikembangkan secara bertahap hingga seluruh data terklasifikasi ke dalam kategori konsumsi yang sesuai. Salah satu studi pada tahun 2024 yang menguji model ini untuk mengelompokkan data insentif karyawan juga melaporkan bahwa akurasi yang dihasilkan tetap stabil meskipun data yang digunakan didominasi oleh variabel numerik[14]. Beberapa penelitian terbaru menyatakan bahwa penggabungan

antara algoritma SVM dan C4.5 dapat meningkatkan efektivitas dalam proses klasifikasi, khususnya untuk menyeleksi calon penerima beasiswa. Dalam studi penerimaan beasiswa, penggunaan kedua algoritma tersebut terbukti memberi prediksi yang lebih objektif dan meringankan beban kerja manual dalam seleksi kandidat[15].

Selain itu, sejumlah penelitian terbaru juga memperlihatkan bahwa algoritma C4.5 mampu memberikan dukungan yang kuat bagi pelaku usaha melalui analisis data historis penjualan. Misalnya, sebuah studi yang mengkaji perilaku pelanggan menunjukkan bahwa C4.5 dapat membentuk pola klasifikasi yang akurat sehingga membantu pelaku bisnis memahami kecenderungan permintaan secara lebih terstruktur[16].

Penelitian lainnya yang meneliti pola penjualan di sektor ritel juga membuktikan bahwa model keputusan berbasis C4.5 mampu menghasilkan aturan prediksi yang mudah diterapkan dalam perencanaan strategi penjualan harian[17].

Temuan serupa diperoleh pada studi penjualan produk makanan, di mana C4.5 digunakan untuk memprediksi volume produksi harian dan memberikan hasil yang stabil berdasarkan pola konsumsi sebelumnya[18].

Temuan-temuan ini menunjukkan bahwa algoritma C4.5 memiliki potensi besar untuk membantu UMKM dalam menentukan jumlah produksi yang lebih efisien berdasarkan data operasional yang mereka miliki.

Kinerja model dievaluasi menggunakan metode 10-fold cross validation dengan menghitung rata-rata hasil dari setiap fold untuk memperoleh gambaran generalisasi yang lebih akurat[19]. Dataset dibagi menjadi sepuluh bagian seimbang; pada tiap iterasi satu bagian digunakan sebagai data uji dan sembilan bagian sebagai data latih. Proses ini diulang sepuluh kali hingga seluruh data bergantian menjadi data uji. Nilai akhir diperoleh dari rata-rata seluruh iterasi, sehingga evaluasi lebih konsisten dan mampu mengurangi bias dibandingkan pembagian data satu kali.

d. Hasil Rekomendasi

Pada tahap akhir, aturan keputusan dari model klasifikasi digunakan sebagai dasar penyusunan rekomendasi produksi berdasarkan tingkat konsumsi harian, sehingga keputusan operasional menjadi lebih terukur dan berbasis data. Pendekatan ini membantu usaha kecil mengurangi risiko kelebihan produksi yang dapat menyebabkan pemborosan makanan. Berbagai studi pada UMKM juga menunjukkan bahwa food waste berdampak tidak hanya pada lingkungan, tetapi juga pada kondisi ekonomi dan efisiensi operasional usaha. Oleh karena itu, pemahaman mengenai pola konsumsi dan penerapan strategi pengelolaan limbah menjadi hal yang penting agar perencanaan produksi dapat dilakukan secara lebih tepat dan efektif[4]. Selain itu, berbagai program edukasi yang berfokus pada penerapan konsep zero-waste di lingkungan UMKM menunjukkan bahwa penyuluhan dan pendampingan yang memanfaatkan data operasional usaha mampu menurunkan jumlah limbah dapur secara signifikan. Upaya semacam ini terbukti membantu pelaku usaha memahami titik-titik pemborosan dan mendorong mereka untuk menerapkan praktik produksi yang lebih efisien[20].

Dalam praktik UMKM, teknik peramalan permintaan semakin dimanfaatkan untuk mendukung penentuan jumlah produksi yang lebih akurat. Melalui pendekatan *forecasting*, pelaku usaha dapat memperkirakan kebutuhan pada periode mendatang sehingga proses produksi dan pengelolaan persediaan dapat dirancang secara lebih terencana dan efisien[21].

2.2 Analisis Data

Dataset terdiri dari data harian selama periode tertentu dengan jumlah record sebanyak (92) data. Atribut yang digunakan terdiri dari (6) atribut input dan 1 atribut kelas. Kategori tingkat konsumsi ditentukan berdasarkan perbandingan antara jumlah penjualan dan jumlah produksi yang kemudian diklasifikasikan menjadi tiga kelas yaitu tinggi, sedang, dan rendah. Berikut adalah contoh analisis data pengkategorian atribut Sumber :Hasil pengolahan data menggunakan RapidMiner (2016).

Tabel 1. Analisis Data Pengkategorian Atribut

Atribut	Indikator	Kategori
Jumlah Produksi	1. >110 porsi	1. Tinggi
	2. 80-100 porsi	2. Sedang
	3. <80 porsi	3. Rendah
Jumlah Penjualan	1. ≥90 porsi	1. Tinggi
	2. 60-89 porsi	2. Sedang
	3. <60 porsi	3. Rendah
Sisa Makanan	1. >30 porsi	1. Tinggi
	2. 10-30 porsi	2. Sedang
	3. <10 porsi	3. Rendah
Jenis Menu	1. Lontong,	1. Menu 1
	2. Nasi Guruh	2. Menu 2
	3. Nasi Sayur	3. Menu 3
	4. Mie	4. Menu 4

Cuaca	1. Cerah 2. Mendung 3. Hujan	1. Tinggi 2. Sedang 3. Rendah
Hari Operasional	1. Weekend (Sabtu-Minggu) 2. Weekday (Senin-Jumat)	1. Weekend 2. Weekday

Tabel 1 diatas merupakan contoh analisis data berdasarkan pengkategorian atribut, dengan 3 kategori konsumsi yaitu tinggi, rendah dan sedang. Pada tahap analisis data, setiap variabel dalam dataset dikonversi ke dalam bentuk kategori untuk menyederhanakan struktur data dan menyesuaikan proses klasifikasi menggunakan algoritma C4.5. Meskipun C4.5 mampu menangani data numerik maupun nominal, proses pengkategorian dilakukan agar pola hubungan antar variabel dapat diidentifikasi dengan lebih jelas dan terstruktur. Sementara itu, variabel non-numerik seperti jenis menu, kondisi cuaca, dan hari operasional dikelompokkan berdasarkan karakteristik yang relevan dengan operasional UMKM. Pengelompokan ini bertujuan untuk memperjelas representasi data sehingga algoritma dapat menentukan atribut yang paling berpengaruh terhadap tingkat konsumsi secara lebih sistematis. Dengan demikian, pengkategorian atribut dalam tabel menjadi dasar yang kuat dalam penyusunan model klasifikasi untuk mengidentifikasi pola signifikan antara produksi, menu, kondisi cuaca, dan tingkat konsumsi harian.

2.3 Data Mining

Data mining pada dasarnya adalah proses untuk menggali pola, hubungan, serta informasi penting dari kumpulan data yang berukuran besar. Proses ini memanfaatkan berbagai pendekatan seperti analisis statistik, metode machine learning, dan algoritma komputasi sehingga data mentah dapat diolah menjadi wawasan yang bermanfaat bagi proses pengambilan keputusan[22]. Melalui teknik ini, peneliti maupun pelaku usaha dapat memahami kecenderungan dan perilaku tertentu yang biasanya tidak terlihat jika hanya mengandalkan pengamatan biasa.

Proses data mining umumnya terdiri dari beberapa langkah penting yang dilakukan secara sistematis. Tahapan tersebut biasanya dimulai dari pemilihan data yang relevan, dilanjutkan dengan proses pembersihan untuk menghilangkan nilai yang salah, tidak lengkap, atau tidak konsisten. Setelah itu, data yang sudah bersih ditransformasikan ke dalam format yang lebih sesuai untuk dianalisis. Pada tahap berikutnya, data diproses menggunakan algoritma tertentu untuk membangun model yang dibutuhkan. Terakhir, hasil model dievaluasi guna memastikan tingkat akurasi serta kelayakan penggunaannya dalam menyelesaikan permasalahan yang diteliti[23]. Rangkaian tahapan ini penting agar informasi yang dihasilkan benar-benar mencerminkan kondisi sebenarnya dan dapat dimanfaatkan secara efektif dalam pengambilan keputusan.

Dalam penelitian ini, teknik klasifikasi digunakan untuk memodelkan hubungan antara variabel operasional UMKM dengan tingkat konsumsi harian guna mengidentifikasi pola yang terbentuk berdasarkan data historis. Model yang dihasilkan mengelompokkan kondisi operasional ke dalam kategori tertentu sebagai dasar dalam mendukung pengambilan keputusan produksi yang lebih terukur. Informasi ini sangat berguna agar strategi produksi dan distribusi dapat direncanakan secara lebih tepat. Salah satu penelitian yang menerapkan algoritma C4.5 menunjukkan bahwa metode klasifikasi berbasis pohon keputusan mampu mengidentifikasi faktor-faktor penentu dalam penyediaan stok misalnya jumlah sisa persediaan setiap minggu sehingga proses pengelolaan stok menjadi lebih efisien dan risiko kelebihan atau kekurangan barang dapat ditekan[24]

2.4 Algoritma C4.5

Algoritma C4.5 merupakan penyempurnaan dari metode ID3 yang sebelumnya dikembangkan oleh beberapa penelitian pada awal tahun 1990-an[25]. Dalam menentukan atribut terbaik, algoritma C4.5 menggunakan ukuran *gain ratio* sehingga proses pembagian data menjadi lebih proporsional serta mengurangi bias terhadap atribut yang memiliki jumlah kategori yang banyak. Selain itu, algoritma ini mampu menangani data yang memiliki nilai hilang (*missing value*), menerapkan proses *pruning* untuk mengurangi kompleksitas pohon keputusan, serta dapat bekerja secara efektif pada data numerik melalui pembentukan nilai ambang (*threshold*). Dengan berbagai keunggulan tersebut, C4.5 mampu menghasilkan struktur pohon keputusan yang lebih stabil dan memiliki tingkat akurasi yang baik, sehingga sesuai diterapkan dalam berbagai kebutuhan klasifikasi di bidang bisnis, termasuk analisis penjualan pada UMKM. Algoritma C4.5 dipilih karena mampu menghasilkan model pohon keputusan yang mudah dipahami serta dapat diturunkan menjadi aturan keputusan (IF-THEN) untuk membantu proses pengambilan keputusan operasional. Penerapan C4.5 pada studi klasifikasi juga menunjukkan bahwa model ini dapat digunakan untuk membangun struktur keputusan yang interpretatif dari data atribut[26].

Berikut ini merupakan tahapan kerja algoritma C4.5 yang dilakukan melalui proses perhitungan :

a. Menyiapkan data latih

Pada tahap awal, seluruh dataset yang digunakan untuk proses pembelajaran dipersiapkan terlebih dahulu agar dapat diolah oleh algoritma

b. Menghitung nilai entropy

Nilai entropy digunakan untuk mengukur tingkat ketidakpastian atau keacakan dalam suatu kumpulan data. Perhitungan entropy dilakukan menggunakan persamaan (1) :

$$\text{Entropy}(\text{total}) = -\left(\frac{B}{A}\right)\log_2\left(\frac{B}{A}\right) - \left(\frac{C}{A}\right)\log_2\left(\frac{C}{A}\right) \quad (1)$$

c. Menghitung nilai gain

Nilai gain dipakai untuk menentukan seberapa besar suatu atribut mampu mengurangi ketidakpastian. Perhitungannya mengikuti persamaan (2) :

$$\text{Gain}(E) = \text{Entropi}(D) - \left(\frac{|H|}{|A|}\right) \cdot \text{Entropi}(K) + \left(\frac{|O|}{|A|}\right) \cdot \text{Entropi}(R) \quad (2)$$

d. Menghitung nilai split information

Split information digunakan untuk menyesuaikan nilai gain agar tidak bias terhadap atribut dengan banya cabang. Rumusnya terdapat pada persamaan (3) :

$$\text{SplitInformation}(F) = -\frac{|H|}{|A|}\log_2\left(\frac{|H|}{|A|}\right) - \left(\frac{|O|}{|A|}\right) \quad (3)$$

e. Menentukan nilai gain ratio

Gain ratio diperoleh dengan membagi nilai gain dengan split information. Atribut yang memiliki gain ratio tertinggi dipilih sebagai akar (*root*) dalam pembentukan pohon keputusan. Rumusnya ditunjukkan pada persamaan (4) :

$$\text{GainRatio}(G) = \frac{E}{F} \quad (4)$$

3. HASIL DAN PEMBAHASAN

3.1 Deskripsi Data Penelitian

Data yang digunakan dalam penelitian ini berupa catatan kegiatan operasional harian pada salah satu UMKM penyedia makanan sarapan yang berlokasi di Kota Medan. Pengumpulan data dilakukan secara langsung oleh pelaku usaha melalui pencatatan rutin selama rentang waktu Juli hingga Agustus 2025, dengan total sampel data yang dianalisis sebanyak 92 entri. Variabel yang tercakup dalam dataset meliputi jumlah produksi makanan, volume penjualan harian, jumlah sisa makanan, jenis menu yang disajikan, kondisi cuaca, serta hari operasional usaha. Seluruh data tersebut dimanfaatkan sebagai dasar analisis dalam membangun model klasifikasi tingkat konsumsi makanan dengan menerapkan algoritma C4.5, di mana penentuan variabel target dilakukan sesuai dengan tujuan penelitian.

Metode pengumpulan data yang digunakan dalam penelitian ini memiliki kesesuaian dengan sejumlah studi sebelumnya yang menjadikan data operasional UMKM sebagai dasar untuk mengkaji faktor-faktor yang berpengaruh terhadap performa usaha. Salah satu penelitian menegaskan bahwa keberadaan UMKM berkontribusi signifikan terhadap perekonomian, khususnya dalam membuka lapangan pekerjaan dan meningkatkan pendapatan masyarakat. Oleh karena itu, pemahaman terhadap perilaku konsumen serta pola penjualan harian menjadi aspek penting dalam penyusunan strategi pemasaran dan pengelolaan operasional usaha[27].

Selain itu, penelitian lain yang mengkaji peran teknologi finansial terhadap perkembangan UMKM di Indonesia dengan memanfaatkan data primer menunjukkan bahwa faktor-faktor operasional, seperti ketersediaan modal dan pemanfaatan teknologi, memiliki pengaruh yang signifikan terhadap kinerja usaha[28].

Pendekatan yang sejalan juga diterapkan dalam kajian mengenai pemasaran digital pada UMKM, yang menyimpulkan bahwa pemanfaatan strategi promosi berbasis digital berperan dalam meningkatkan volume penjualan produk di tengah perkembangan era digital. Temuan tersebut dipublikasikan dalam jurnal pengabdian masyarakat yang berfokus pada aspek pemberdayaan dan inovasi, serta dapat diakses melalui repositori jurnal yang bersifat terbuka[29].

Berdasarkan kajian-kajian sebelumnya, penelitian ini menggunakan data operasional harian UMKM makanan sebagai dasar untuk mengidentifikasi pola konsumsi serta fluktuasi produksi yang dipengaruhi oleh berbagai faktor, baik dari sisi internal maupun eksternal usaha. Data yang telah dikumpulkan selanjutnya diproses melalui beberapa tahapan preprocessing, meliputi pembersihan data, pengubahan data ke dalam bentuk kategorikal, serta proses pengkodean. Tahapan tersebut dilakukan untuk memastikan data siap dianalisis menggunakan algoritma C4.5 dalam membangun model klasifikasi tingkat konsumsi makanan.

3.2 Tahap Preprocessing Data

Dalam kerangka data mining, tahap preprocessing diharapkan mampu memastikan bahwa data yang digunakan untuk proses klasifikasi memenuhi persyaratan kualitas yang dibutuhkan oleh model analitik. Data idealnya berada dalam kondisi bersih, konsisten, dan relevan dengan tujuan pemodelan. Proses ini meliputi kegiatan pembersihan data, integrasi data, transformasi data, serta pemilihan atribut, karena setiap tahapan tersebut berperan dalam menentukan kualitas dan kinerja model klasifikasi yang dihasilkan[30].

Secara empiris, tahapan preprocessing pada penelitian ini telah dilaksanakan dan dijelaskan secara rinci. Hasil preprocessing menunjukkan bahwa seluruh data tidak mengandung nilai kosong (*missing value*), telah mengalami

penyeragaman format pada atribut kategorikal, serta telah diseleksi berdasarkan kesesuaiannya dengan tujuan penelitian. Kondisi ini menunjukkan bahwa data telah memenuhi kriteria kesiapan untuk tahap pemodelan.

Sejumlah penelitian terdahulu menegaskan bahwa keberhasilan model klasifikasi sangat dipengaruhi oleh kualitas preprocessing data, khususnya pada metode klasifikasi berbasis pohon keputusan seperti Decision Tree C4.5[30]. Meskipun demikian, masih terbatas kajian yang secara eksplisit menekankan keterkaitan antara kesiapan atribut kategorikal hasil preprocessing dengan pembentukan model C4.5 pada konteks penelitian tertentu, sehingga terdapat kesenjangan penelitian pada aspek tersebut.

Berdasarkan kondisi tersebut, penelitian ini menjadi relevan untuk dilakukan guna memastikan bahwa data hasil preprocessing telah memenuhi kebutuhan teknis dalam pembentukan model klasifikasi. Tujuan utama penelitian ini adalah memanfaatkan data yang telah terstruktur secara tepat untuk membangun model klasifikasi menggunakan algoritma Decision Tree C4.5 sesuai dengan karakteristik data dan tujuan analisis yang telah ditetapkan.

3.3 Pembentukan Variabel Target

Distribusi kelas pada variabel target dilakukan untuk mengetahui proporsi masing-masing kategori tingkat konsumsi dalam dataset. Analisis ini penting untuk memastikan bahwa data yang digunakan dalam proses klasifikasi tidak mengalami ketidakseimbangan kelas (*class imbalance*) yang dapat memengaruhi performa model.

Berdasarkan hasil pengolahan terhadap 92 data operasional harian UMKM, diperoleh distribusi tingkat konsumsi sebagai berikut:

Tabel 2. Distribusi Kelas Tingkat Konsumsi

No	Tingkat Konsumsi	Jumlah Data	Persentase (%)
1	Sedang	56	60,87 %
2	Rendah	20	21,74 %
3	Tinggi	16	17,39 %
	Total	92	100 %

Sumber : Hasil pengolahan data menggunakan RapidMiner (2016)

Tabel 2 menyajikan hasil performa dari distribusi tingkat konsumsi yang dikelompokkan menjadi 3 bagian yaitu sedang, rendah dan tinggi. Distribusi tersebut digunakan untuk mengetahui proporsi tingkat konsumsi dalam data set. Berdasarkan hasil distribusi data menunjukkan bahwa kategori tingkat konsumsi sedang mendominasi dataset dengan proporsi sebesar 60,87%. Sementara itu, kategori tinggi dan rendah memiliki proporsi yang lebih kecil namun masih dalam rentang yang wajar. Kondisi ini menunjukkan bahwa dataset cenderung memiliki distribusi kelas yang cukup seimbang meskipun terdapat dominasi pada kelas sedang. Dengan demikian, model klasifikasi C4.5 tetap dapat dibangun tanpa memerlukan teknik penanganan khusus terhadap ketidakseimbangan data.

3.4 Perhitungan Entropy Awal

Sebelum menentukan atribut terbaik dalam pembentukan pohon keputusan, terlebih dahulu dihitung nilai entropy total dataset. Entropy digunakan untuk mengukur tingkat ketidakpastian distribusi kelas pada data awal.

Berdasarkan distribusi kelas pada Tabel 2, dari total 92 data diperoleh:

- Tinggi = 16 data
- Sedang = 56 data
- Rendah = 20 data

Sehingga entropy total dihitung menggunakan rumus:

$$Entropy(S) = - \sum_{i=0}^n p_i \log_2 p_i$$

Substitusi nilai probabilitas masing-masing kelas:

$$Entropy(S) = - \left(\frac{16}{92} \log_2 \frac{16}{92} \right) - \left(\frac{56}{92} \log_2 \frac{56}{92} \right) - \left(\frac{20}{92} \log_2 \frac{20}{92} \right)$$

Perhitungan :

$$\begin{aligned} &= -(0,1739 \log_2 0,1739) - (0,6087 \log_2 0,6087) - (0,2174 \log_2 0,2174) \\ &= -(0,1739 \times -2,523) - (0,6087 \times -0,715) - (0,2174 \times -2,201) \\ &= -(0,1739 \times -2,523) - (0,6087 \times -0,715) - (0,2174 \times -2,201) \\ &= 0,438 + 0,435 + 0,478 = 0,438 + 0,435 + 0,478 \\ Entropy(S) &= 1,351 \end{aligned}$$

Nilai entropy total sebesar 1,351 menunjukkan bahwa dataset memiliki tingkat ketidakpastian yang cukup tinggi karena distribusi kelas masih bercampur antara kategori tinggi, sedang, dan rendah. Berdasarkan metode perhitungan entropy dan gain yang telah dijelaskan, selanjutnya dilakukan perhitungan terhadap dataset penelitian untuk menentukan atribut terbaik dalam pembentukan pohon keputusan.

Tabel 3. Hasil Perhitungan Algoritma C4.5

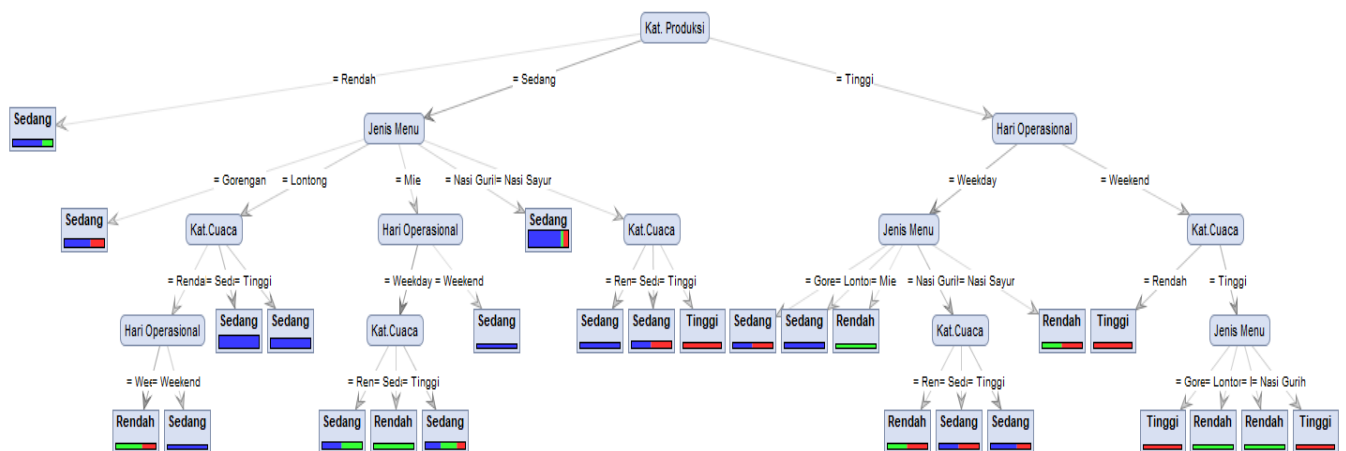
Atribut	Kategori	Jumlah(S)	Tinggi	Sedang	Rendah	Entropy	Gain	Split Info	Gain Ratio
Jenis Menu	Gorengan	9	4	5	0	0.9911	0.2152	2.2062	0.0976
	Lontong	27	3	21	3	0.9864			
	Mie	18	1	7	10	1.2327			
	Nasi Gurih	26	7	17	2	1.1951			
	Nasi Sayur	12	5	6	1	1.3250			
Hari Operasional	Weekday	65	13	38	14	1.3942	0.0241	0.8732	0.0277
	Weekend	27	7	18	2	1.1730			
Kat.Cuaca	Rendah	25	6	13	6	1.4788	0.0604	1.5665	0.0385
	Sedang	30	3	24	3	0.9219			
	Tinggi	37	11	19	7	1.4685			
Kat. Produksi	Rendah	32	3	18	11	1.3085	0.1136	1.0163	0.1117
	Sedang	43	14	28	1	0.9673			
	Tinggi	17	3	10	4	1.4466			

Sumber : Hasil pengolahan data menggunakan RapidMiner (2016)

Tabel 3 diatas merupakan hasil pengolahan data set menggunakan algoritma data C4.5 untuk mengetahui nilai entropy dan gain ratio. Berdasarkan hasil perhitungan entropy dan gain ratio, atribut kategori produksi memiliki nilai gain ratio terbesar dibandingkan atribut lainnya. Nilai gain ratio yang lebih tinggi menunjukkan bahwa atribut tersebut paling efektif dalam mengurangi ketidakpastian data dan memisahkan kelas konsumsi secara optimal. Dengan demikian, Kategori Produksi dipilih sebagai akar (root) dalam pembentukan pohon keputusan karena memiliki kemampuan paling signifikan dalam menjelaskan variasi tingkat konsumsi harian.

3.5 Pembentukan Pohon Keputusan

Berdasarkan perhitungan nilai entropy, information gain, dan gain ratio untuk seluruh atribut yang dianalisis, diketahui bahwa atribut kategori produksi memiliki nilai gain ratio paling tinggi dibandingkan atribut lainnya. Hal ini mengindikasikan bahwa Kategori Produksi paling efektif dalam menurunkan tingkat ketidakpastian data serta paling baik dalam membedakan kelas tingkat konsumsi harian. Oleh karena itu, atribut tersebut ditetapkan sebagai akar (root) pada pembentukan pohon keputusan dengan algoritma C4.5. Selanjutnya, proses pembentukan cabang dilakukan secara bertahap dengan memilih atribut yang memiliki nilai gain ratio tertinggi pada masing-masing subset data, hingga terbentuk struktur klasifikasi yang paling optimal.



Gambar 2. Struktur Pohon Keputusan

Gambar 2 diatas merupakan struktur pohon keputusan menunjukkan bahwa kategori produksi merupakan atribut utama dalam menentukan tingkat konsumsi harian. Pada produksi rendah, tingkat konsumsi cenderung berada pada kategori Sedang. Pada produksi sedang, pemisahan dipengaruhi oleh jenis menu dan dalam kondisi tertentu dilanjutkan

oleh kategori cuaca serta hari operasional. Sementara pada produksi tinggi, hari operasional dan cuaca menjadi faktor pembeda utama.

Secara keseluruhan, model menunjukkan bahwa jumlah produksi merupakan faktor paling dominan, sedangkan jenis menu, cuaca, dan hari operasional berperan sebagai faktor pendukung dalam menjelaskan variasi tingkat konsumsi. Struktur ini kemudian menjadi dasar dalam penyusunan rekomendasi produksi untuk mendukung pengendalian food waste pada UMKM sarapan pagi.

Setiap jalur dari akar hingga simpul daun pada pohon keputusan menghasilkan aturan keputusan (rules). Aturan tersebut menggambarkan hubungan antara atribut operasional dengan tingkat konsumsi harian. Rules yang terbentuk dari pohon keputusan ditunjukkan pada Tabel 4.

Tabel 4. Aturan Keputusan (Decision Rules) Algoritma C4.5

No	Rules	Keputusan
1	IF Kat.Produksi = Rendah	Sedang
2	IF Kat.Produksi = Sedang AND Jenis Menu = Nasi Gurih	Sedang
3	IF Kat.Produksi = Sedang AND Jenis Menu = Nasi Sayur	Sedang
4	IF Kat.Produksi = Sedang AND Jenis Menu = Gorengan AND Kat.Cuaca	Rendah
5	IF Kat.Produksi = Sedang AND Jenis Menu = Gorengan AND Kat.Cuaca = Sedang	Sedang
6	IF Kat.Produksi = Sedang AND Jenis Menu = Gorengan AND Kat.Cuaca = Tinggi	Sedang
7	IF Kat.Produksi = Sedang AND Jenis Menu = Mie AND Hari Operasional = Weekend	Sedang
8	IF Kat.Produksi = Sedang AND Jenis Menu = Mie AND Hari Operasional = Weekday AND Kat.Cuaca = Rendah	Sedang
9	IF Kat.Produksi = Sedang AND Jenis Menu = Mie AND Hari Operasional = Weekday AND Kat.Cuaca = Sedang	Rendah
10	IF Kat.Produksi = Sedang AND Jenis Menu = Mie AND Hari Operasional = Weekday AND Kat.Cuaca = Tinggi	Sedang
11	IF Kat.Produksi = Tinggi AND Hari Operasional = Weekday AND Jenis Menu = Gorengan	Sedang
12	IF Kat.Produksi = Tinggi AND Hari Operasional = Weekday AND Jenis Menu = Lontong	Rendah

Sumber : Hasil pengolahan data menggunakan RapidMiner (2016)

Berdasarkan tabel 4 diatas, aturan keputusan tersebut dapat digunakan sebagai dasar dalam menentukan rekomendasi produksi harian. Dengan memanfaatkan pola hubungan antara jumlah produksi, jenis menu, kondisi cuaca, dan hari operasional, pelaku UMKM dapat menyesuaikan jumlah produksi agar lebih sesuai dengan tingkat konsumsi sehingga potensi terjadinya food waste dapat diminimalkan.

Secara lebih rinci, hasil interpretasi menunjukkan bahwa kategori produksi menjadi faktor dominan dalam menentukan tingkat konsumsi, di mana sebagian besar kondisi menghasilkan keputusan pada kategori konsumsi *sedang*. Hal ini mengindikasikan bahwa permintaan pasar cenderung stabil pada tingkat menengah, sehingga produksi yang terlalu tinggi berpotensi meningkatkan sisa makanan. Selain itu, jenis menu juga berpengaruh terhadap variasi konsumsi. Menu seperti nasi gurih dan nasi sayur cenderung konsisten menghasilkan tingkat konsumsi sedang, yang menunjukkan bahwa kedua menu tersebut memiliki permintaan yang relatif stabil.

Sementara itu, pada menu gorengan dan mie, terlihat bahwa faktor eksternal seperti kondisi cuaca dan hari operasional turut memengaruhi tingkat konsumsi. Misalnya, pada kondisi tertentu (seperti cuaca sedang atau hari kerja), konsumsi dapat menurun menjadi rendah, yang mengindikasikan perlunya penyesuaian produksi untuk menghindari kelebihan stok. Selain itu, pada kategori produksi tinggi, beberapa kombinasi kondisi justru menghasilkan konsumsi rendah atau sedang, yang menunjukkan bahwa peningkatan produksi tidak selalu sejalan dengan peningkatan permintaan.

Dengan demikian, aturan keputusan ini tidak hanya berfungsi sebagai alat prediksi, tetapi juga memberikan wawasan praktis mengenai pola konsumsi pelanggan. Pelaku UMKM dapat menggunakan informasi ini untuk merencanakan produksi secara lebih adaptif berdasarkan kombinasi faktor operasional yang ada, sehingga efisiensi produksi meningkat dan risiko *food waste* dapat ditekan secara signifikan

3.6 Evaluasi Kinerja Model

Evaluasi kinerja model dilakukan untuk mengetahui kemampuan algoritma C4.5 dalam mengklasifikasikan tingkat konsumsi harian berdasarkan atribut operasional yang digunakan dalam penelitian. Proses evaluasi pada penelitian ini menggunakan metode 10-fold cross validation yang tersedia pada perangkat lunak RapidMiner. Metode ini dipilih karena mampu memberikan estimasi performa model yang lebih stabil dengan cara membagi dataset menjadi sepuluh bagian dengan proporsi yang relatif sama.

Pada setiap iterasi, sembilan bagian data digunakan sebagai data latih (training data) dan satu bagian lainnya digunakan sebagai data uji (testing data). Proses ini dilakukan sebanyak sepuluh kali sehingga setiap bagian data akan digunakan sebagai data uji secara bergantian. Nilai performa model kemudian diperoleh dari rata-rata hasil pengujian pada seluruh iterasi sehingga dapat memberikan gambaran kemampuan generalisasi model yang lebih akurat.

Berdasarkan hasil pengujian menggunakan metode tersebut, model klasifikasi C4.5 memperoleh nilai accuracy sebesar 57,56% dengan deviasi sebesar $\pm 7,85\%$. Nilai tersebut menunjukkan bahwa model mampu mengklasifikasikan tingkat konsumsi harian dengan tingkat ketepatan sebesar 57,56%.

Untuk melihat performa klasifikasi secara lebih rinci, hasil pengujian juga disajikan dalam bentuk confusion matrix yang menggambarkan perbandingan antara kelas aktual dengan hasil prediksi model.

Tabel 5. Confusion Matrix Hasil Klasifikasi

Prediksi / Aktual	Sedang	Rendah	Tinggi
Sedang	44	8	11
Rendah	6	4	4
Tinggi	6	4	5

Sumber : Hasil pengolahan data menggunakan RapidMiner (2016)

Tabel 5 diatas merupakan bentuk tabel confusion matrix digunakan untuk membandingkan nilai kelas actual dengan hasil prediksi model. Berdasarkan hasil confusion matrix tersebut, dapat dilihat bahwa model memiliki kemampuan klasifikasi yang paling baik pada kategori Sedang, dengan nilai recall sebesar 78,57%. Hal ini menunjukkan bahwa sebagian besar data dengan kategori konsumsi sedang berhasil diprediksi dengan benar oleh model. Sementara itu, performa model pada kategori Rendah dan Tinggi masih relatif lebih rendah, yang menunjukkan bahwa beberapa data pada kategori tersebut masih mengalami kesalahan klasifikasi.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa algoritma C4.5 mampu mengidentifikasi pola hubungan antara jumlah produksi, jenis menu, kondisi cuaca, serta hari operasional terhadap tingkat konsumsi harian. Meskipun tingkat akurasi model belum mencapai nilai yang sangat tinggi, model yang dihasilkan tetap dapat memberikan gambaran pola konsumsi yang bermanfaat bagi pelaku UMKM dalam menentukan jumlah produksi yang lebih sesuai dengan permintaan pelanggan sehingga potensi terjadinya food waste dapat diminimalkan.

4. KESIMPULAN

Berdasarkan hasil perhitungan entropy, information gain, dan gain ratio, diketahui bahwa atribut kategori produksi memiliki nilai gain ratio terbesar sehingga dipilih sebagai akar dalam pembentukan pohon keputusan. Struktur pohon keputusan yang dihasilkan menunjukkan bahwa jumlah produksi merupakan faktor utama yang memengaruhi tingkat konsumsi harian, sedangkan atribut lain seperti jenis menu, kondisi cuaca, dan hari operasional berperan sebagai faktor pendukung dalam proses klasifikasi. Evaluasi model menggunakan metode 10-fold cross validation menunjukkan bahwa algoritma C4.5 memperoleh nilai accuracy sebesar 57,56%. Hasil ini menunjukkan bahwa model mampu mengidentifikasi pola hubungan antara variabel operasional UMKM dengan tingkat konsumsi harian. Aturan keputusan yang dihasilkan dari pohon keputusan dapat digunakan sebagai dasar dalam menentukan rekomendasi jumlah produksi sehingga pelaku UMKM dapat menyesuaikan jumlah produksi dengan pola konsumsi pelanggan. Dengan demikian, model klasifikasi yang dihasilkan berpotensi membantu pelaku usaha dalam mengurangi sisa makanan (food waste) serta meningkatkan efisiensi produksi. Penelitian selanjutnya diharapkan dapat menggunakan dataset yang lebih besar serta menambahkan atribut lain agar performa model klasifikasi dapat ditingkatkan.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada pihak UMKM sarapan pagi di Kota Medan yang telah bersedia memberikan data operasional harian sebagai bahan penelitian. Penulis juga menyampaikan apresiasi kepada dosen pembimbing dan pihak Program Studi yang telah memberikan arahan, masukan, serta dukungan selama proses penyusunan penelitian ini. Selain itu, penulis berterima kasih kepada semua pihak yang turut membantu dalam pengumpulan data, proses pengolahan data, hingga penyusunan artikel ini sehingga penelitian dapat diselesaikan dengan baik.

REFERENCE

- [1] E. Adinda Yestina *et al.*, "Implementasi Algoritma C4.5 pada Analisis Faktor Risiko Penyakit Jantung Koroner," *Jurnal Riset Komputer*, vol. 12, no. 5, pp. 2407–389, 2025, doi: 10.30865/jurikom.v12i5.8852.
- [2] M. Rudi Fanani and E. Bendra Agustina, "Smart Comp: Jurnalnya Orang Pintar Komputer Klasifikasi Strategi Penjualan Produk UMKM dengan Penerapan fitur seleksi Forward Selection pada Algoritma C4.5," vol. 14, doi: 10.30591/smartcomp.v14i4.6842.
- [3] W. Kustiawan, I. Safitri, and R. Q. Arsyad, "STRATEGI ADIPTIF UMKM KULINER TRADISIONAL MEDAN DALAM MENGHADAPI DOMINASI WARALABA ASING DI ERA GLOBALISASI," May 2025.
- [4] L. Yuliana, "Potensi Gerakan Anti Food Waste dalam Penguatan Perekonomian UMKM," *Efektor*, vol. 9, no. 2, pp. 286–295, Dec. 2022, doi: 10.29407/e.v9i2.18647.
- [5] L. Lybaws, Y. F. Baliwati, and I. Tanzaha, "Determinants of Food Waste Among Household in Rural and Urban Areas of Bogor Regency," *Amerta Nutrition*, vol. 9, no. 1, pp. 1–13, Mar. 2025, doi: 10.20473/amnt.v9i1.2025.1-13.
- [6] T. N. Fajri and E. R. Shauki, "POTENSI FOOD LOSS DAN FOOD WASTE PADA UMKM: MFCA, NUDGING DAN NEUTRALIZATION THEORY," *Jurnal Aplikasi Akuntansi*, vol. 7, no. 2, pp. 328–345, Feb. 2023, doi: 10.29303/jaa.v7i2.187.

- [7] B. Dwi Saputra, “PENGARUH BIG DATA ANALYTICS MENGGUNAKAN ALGORITMA C4.5, STRATEGIC DECISION, DAN MARKETING PERFORMANCE SERTA DAMPAKNYA TERHADAP INTENTION TO USE PADA UMKM DI DKI JAKARTA,” vol. 23, no. 11, 2025, doi: 10.8734/mmnae.v1i2.359.
- [8] M. I. P. N. Pramuja Angkasa, “Penerapan Business Intelligence dalam Pengambilan Keputusan Bisnis: Studi Pendekatan Praktis untuk UMKM di Indonesia,” *Jurnal Penelitian Ilmu-Ilmu Sosial*, vol. 2, no. 11, Jun. 2025.
- [9] L. D. Hasan *et al.*, “Social, Humanities, and Educational Studies SHES: Conference Series 7 (4) (2024) 122-132 Kampanye ‘Food Waste’ pada UMKM Kuliner di Kota Makassar.” [Online]. Available: <https://jurnal.uns.ac.id/shes>
- [10] S. M. R. N. A. Daniella C. Sampepajung, “ANALISIS CARBON FOOTPRINT PADA UMKM MAKANAN DAN MINUMAN MELALUI INTEGRASI LEAN DAN GREEN WASTE PRODUCTION,” *Jurnal Bisnis, Manajemen, dan Informatika (JBMI)*, vol. 19, no. 3, Feb. 2023.
- [11] S. Nadiroh, S. Prawitasari, and N. Fathiyah Fauzi, “Potensi Kehilangan Ekonomi akibat Food Loss dan Food Waste di Kalangan Mahasiswa Universitas Muhammadiyah Jember,” *Berkala Ilmiah Pertanian*, vol. 8, no. 3, Aug. 2025, doi: 10.19184/bip.v8i3.53704.
- [12] A. Djaini, A. Haslinah, and H. N. Muthmainah, “Strategi Pengurangan Sampah Makanan Menggunakan Analisis Data dan Teknologi,” Oct. 2023.
- [13] R. Fathurrohman, M. Fajri S, S. M. Cahyono, D. F. Abdillah, and V. Ramadhan, “Data Mining Pada Klasifikasi Jamur Menggunakan Algoritma C.45 Berdasarkan Karakteristik Morfologi Mushroom,” Jun. 2024. [Online]. Available: <http://jurnal.bsi.ac.id/index.php/simpatik>
- [14] P. Widia Sari, G. Giwangga, N. Dian Fitriyanti, A. Arlisdia Putra, and H. Irawan, “Implementasi Algoritma C4.5 Untuk Klasifikasi Data Insentif Karyawan PT. Adhi Cakra Utama Mulia,” *Jurnal Komputer Antartika*, vol. 2, no. 3, pp. 95–101, 2024, [Online]. Available: <https://ejournal.mediaantartika.id/index.php/jka>
- [15] Hafid Azis Supahri, Rahmadden, Syarfi Aziz, Zairi Saputra, and Ryan Ismanizan, “Implementasi Algoritma SVM dan C4.5 dalam Klasifikasi Calon Penerimaan Beasiswa,” *Jurnal KomtekInfo*, pp. 81–88, Aug. 2024, doi: 10.35134/komtekinfo.v11i3.524.
- [16] Z. N. Saputra and Z. Fatah, “Pengunaan Data Mining Untuk Mengidentifikasi Pelanggan Beresiko Tinggi Dalam Penjualan Menggunakan Algoritma Decision Tree C4.5,” *JUSIFOR : Jurnal Sistem Informasi dan Informatika*, vol. 4, no. 1, pp. 68–74, Jun. 2025, doi: 10.70609/jusifor.v4i1.5942.
- [17] S. Aditya Narendra *et al.*, “Analisa Penjualan Rokok Dengan Metode Klasifikasi Menggunakan Metode Algoritma C4.5 Pada CV Jaya Berkah Mas,” *Jurnal INSAN (Journal of Information Systems Management Innovation)*, vol. 4, no. 2, 2024, [Online]. Available: <http://jurnal.bsi.ac.id/index.php/jinsan>
- [18] A. B. Prasetyo and I. H. Iksari, “Penerapan Data Mining Untuk Memprediksi Penjualan Kue Brownis Menggunakan Algoritma C4.5 Pada Toko Amanda Brownis,” *Jurnal Ilmu Komputer dan Pendidikan*, vol. 2, no. 1, 2023, [Online]. Available: <https://journal.mediapublikasi.id/index.php/logic>
- [19] R. Tuntun, K. Kusri, and K. Kusnawi, “Analisis Perbandingan Kinerja Algoritma Klasifikasi dengan Menggunakan Metode K-Fold Cross Validation,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 4, p. 2111, Oct. 2022, doi: 10.30865/mib.v6i4.4681.
- [20] S. Suryanti, S. Rudiyan, C. A’in, N. Aulia, and A. E. P. Alfizhari, “Program Edukasi Zero Waste pada Limbah Dapur dan Food Safety dalam Proses Produksi UMKM,” *Jurnal Pengabdian Masyarakat Inovasi Indonesia*, vol. 1, no. 1, pp. 41–44, Dec. 2023, doi: 10.54082/jpmii.269.
- [21] E. Donneli, “Big Data dan Peramalan Permintaan Rantai Pasokan dalam Pemasaran Digital UMKM,” vol. 7, 2024, [Online]. Available: <https://jurnal.untan.ac.id/index.php/MBIC/index>
- [22] A. H. Pradhana, A. A. Firmansyah, and B. M. Islami, “Pemodelan Prediksi Penjualan dan Persediaan dengan RapidMiner beserta Pengelompokan Kategori untuk Mempermudah Perencanaan Stok dan Pengambilan Keputusan Bisnis,” *JAMI: Jurnal Ahli Muda Indonesia*, vol. 6, no. 1, pp. 9–17, Jun. 2025, doi: 10.46510/jami.v6i1.323.
- [23] M. B. Pratama and F. I. Sanjaya, “Prediksi Omset Penjualan Toko Mbak Ning Pasar Cenderamata Menggunakan Long Short Term Memory,” *SIMKOM*, vol. 10, no. 1, pp. 160–167, Jan. 2025, doi: 10.51717/simkom.v10i1.580.
- [24] S. Hartawan *et al.*, “PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA C4. 5 UNTUK MEMPREDIKSI PERSEDIAAN STOK BERAS,” vol. 4, no. 2, 2024, doi: 10.46880/tamika.Vol4No2(SEMNASTIK).pp114-121.
- [25] Y. Fajar Saputra, “Algoritma C4.5 sebagai Penerapan Decision Tree-Based Classification Model untuk Mengklasifikasikan Tingkat Omzet UMKM Berdasarkan Profil Bisnis,” Sep. 2024.
- [26] B. Q. Husaini and J. Jemakmun, “Penerapan Algoritma Decision Tree C45 untuk Klasifikasi Penjurusan Siswa,” *Jurnal Teknologi Informatika dan Komputer*, vol. 9, no. 1, pp. 455–470, Mar. 2023, doi: 10.37012/jtik.v9i1.1512.
- [27] Fatimah Hafni Simanjuntak and Nadra Syahira Putri, “Membangun Ekonomi Berkelanjutan: Kontribusi UMKM dalam Peningkatan Pendapatan Masyarakat,” *Jurnal Nuansa : Publikasi Ilmu Manajemen dan Ekonomi Syariah*, vol. 3, no. 3, pp. 223–230, Jun. 2025, doi: 10.61132/nuansa.v3i3.1955.
- [28] R. N. Linda, “Kontribusi Fintech atas Perkembangan Fan Realita UMKM di Indonesia,” 2025. [Online]. Available: <https://jsr.lib.ums.ac.id/index.php/determinasi?page184>
- [29] M. Rachman Hakim, K. Abdillah, G. Asoloan Sitingjak, A. Anggarani Winadi Prasetyoningtyas, U. Esa Unggul, and J. Arjuna Utara Tol Tomang Kebon Jeruk, “STRATEGI PEMASARAN DIGITAL UMKM KULINER DI ERA MEDIA SOSIAL,” vol. 5, no. 5, 2025, doi: 10.59818/jpm.v5i5.2058.
- [30] T. Gori, A. Sunyoto, and H. Al Fatta, “Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 1, pp. 215–224, Feb. 2024, doi: 10.25126/jtiik.20241118074.