

Implementasi Algoritma C4.5 Untuk Klasifikasi Menu Favorit Pelanggan Pada UMKM Warung Mie Aceh Seafood

Helma Tiara Syifa*, Dicky Nofriansyah, Fifi Sonata

Fakultas Informatika dan Komputer, Sistem Informasi, STMIK Triguna Dharma, Medan, Indonesia
Email: ¹*helmatiarasyifa19@gmail.com, ²nofriansyahdicky@gmail.com, ³fifinsonata2012@gmail.com
Email Penulis Korespondensi: helmatiarasyifa19@gmail.com
Submitted 15-04-2026; Accepted 02-06-2026; Published 30-06-2026

Abstrak

Sektor kuliner yang berkembang pesat menuntut para pelaku Usaha Mikro Kecil dan Menengah (UMKM) untuk menerapkan strategi operasional yang tepat sasaran. Masalah utama yang sering dihadapi oleh manajemen UMKM Warung Mie Aceh Seafood adalah ketidakmampuan dalam memprediksi tingkat kelarisan menu berdasarkan faktor eksternal secara akurat. Hal ini secara langsung berdampak pada inefisiensi pengelolaan persediaan bahan baku basah (*seafood*) akibat sifatnya yang mudah busuk dan kedaluwarsa. Oleh karena itu, penelitian ini bertujuan untuk mengimplementasikan teknologi data mining menggunakan algoritma C4.5 guna mengklasifikasikan menu favorit pelanggan secara prediktif. Sebagai solusi dari permasalahan tersebut, penelitian ini memproses 586 rekam data historis transaksi penjualan pada periode Juli hingga September 2025. Berbeda dengan studi terdahulu, kebaruan penelitian ini terletak pada pengintegrasian variabel lingkungan eksternal, yakni kondisi cuaca dan tipe hari, bersama atribut jenis menu ke dalam satu model pohon keputusan multidimensional yang dievaluasi menggunakan metode 10-fold cross validation. Hasil pengujian menunjukkan bahwa model klasifikasi algoritma C4.5 terbukti sangat prima dengan meraih tingkat akurasi 96,08% serta nilai *class precision* mencapai 100% untuk kelas "Sangat Favorit" dan "Tidak Favorit". Secara praktis, ekstraksi aturan (*rules*) dari model ini berkontribusi spesifik sebagai sistem pendukung keputusan; misalnya memandu manajemen mengurangi stok minuman dingin saat cuaca hujan lebat, atau memaksimalkan persediaan menu 'Jajanan' di akhir pekan, sehingga UMKM dapat menghindari kerugian material sekaligus menekan risiko hilangnya potensi pendapatan secara efisien.

Kata Kunci: Algoritma C4.5; Data Mining; Klasifikasi; Menu Kuliner; Pohon Keputusan

Abstract

The rapidly growing culinary sector requires Micro, Small, and Medium Enterprises (MSMEs) to implement targeted operational strategies. The main problem often faced by the management of the Warung Mie Aceh Seafood MSME is the inability to accurately predict the level of menu popularity based on external factors. This directly impacts the inefficiency of managing wet raw material (*seafood*) inventory due to its perishable nature. Therefore, this study aims to implement data mining technology using the C4.5 algorithm to predictively classify customers' favorite menus. As a solution to these problems, this study processed 586 historical sales transaction data records from July to September 2025. Unlike previous studies, the novelty of this research lies in integrating external environmental variables, namely weather conditions and day types, along with menu type attributes into a single multidimensional decision tree model evaluated using the 10-fold cross-validation method. The test results show that the C4.5 algorithm classification model has proven to be excellent, achieving an accuracy rate of 96.08% and a class precision value of 100% for the "Very Favorite" and "Not Favorite" classes. Practically, extracting rules from this model provides a specific contribution as a decision support system; for instance, guiding management to reduce cold beverage stocks during heavy rain, or maximize the 'Snacks' menu inventory on weekends, enabling the MSME to avoid material losses while efficiently minimizing the risk of lost potential revenue.

Keywords: C4.5 Algorithm; Data Mining; Classification; Culinary Menu; Decision Tree

1. PENDAHULUAN

Sektor kuliner Indonesia berkembang pesat, memaksa UMKM untuk mengembangkan dan menerapkan strategi bisnis khusus agar dapat bersaing. Warung Mie Aceh Seafood adalah UMKM yang menjanjikan. Dalam menjalankan operasional harian, pengelolaan persediaan bahan baku menjadi salah satu tantangan terbesar, terutama karena bahan baku *seafood* seperti cumi dan udang memiliki sifat mudah busuk (*perishable*) dan membutuhkan manajemen stok yang akurat [1]. Pengelolaan dan evaluasi menu merupakan salah satu aspek paling fundamental bagi kelangsungan bisnis di industri kuliner, di mana menu yang dirancang dan dianalisis dengan baik dapat menghasilkan keuntungan yang maksimal bagi UMKM [2]. Berdasarkan rekam jejak operasionalnya, warung ini telah mengumpulkan data transaksi penjualan historis selama tiga bulan terakhir, yaitu dari bulan Juli hingga September 2025. Data transaksi tersebut memuat berbagai atribut yang sangat berharga, seperti Tanggal, Nama Menu, Jenis Menu (seperti Mie Aceh Seafood, Nasi Goreng, Mie Instan, dan Minuman), Harga, Jumlah Terjual, Total Penjualan, Jenis Pesanan (Bungkus atau Makan di tempat), kondisi Cuaca (Cerah atau Hujan), serta tipe Hari (*Weekday* atau *Weekend*).

Meskipun UMKM Warung Mie Aceh Seafood telah memiliki tumpukan data transaksi penjualan harian yang mendetail, data tersebut sejauh ini hanya digunakan sebagai arsip keuangan dan belum diekstraksi menjadi informasi strategis. Permasalahan utama yang dihadapi oleh pengelola adalah ketidakmampuan dalam memprediksi atau mengklasifikasikan menu apa yang menjadi favorit pelanggan pada kondisi-kondisi tertentu, misalnya saat cuaca hujan di akhir pekan (*weekend*) atau saat cuaca cerah di hari kerja (*weekday*). Hal ini mendesak untuk diselesaikan karena permintaan di sektor makanan dan minuman (F&B) pada dasarnya sangat siklikal dan dipengaruhi secara kuat oleh faktor eksternal seperti kondisi cuaca harian dan tipe hari operasional [3] dan sering kali berdampak pada pengambilan keputusan yang tidak optimal, seperti kekurangan bahan baku untuk menu yang tiba-tiba permintaannya melonjak, atau sebaliknya, penumpukan bahan baku untuk menu yang kurang diminati, yang berujung pada kerugian finansial akibat

bahan baku yang kedaluwarsa. Seperti penelitian [4] menerapkan teknik data mining menggunakan *Decision Tree* pada data penjualan kafe lokal untuk menentukan strategi *bundling* produk, yang berhasil meningkatkan efisiensi stok bahan baku. Oleh karena itu, diperlukan sebuah pendekatan analitis untuk menggali pola tersembunyi (*knowledge discovery*) dari data transaksi yang ada, di mana salah satu solusi yang sangat direkomendasikan adalah penerapan teknologi data mining.

Algoritma C4.5 disajikan untuk menyelesaikan masalah kategorisasi menu favorit UMKM. Metode klasifikasi penambangan data seperti Algoritma C4.5 menciptakan model pohon keputusan [5]. Algoritma C4.5 unggul dalam menangani data atribut numerik dan kategorikal [6], termasuk Cuaca, Jenis Pesanan, dan Hari, dan menangani nilai yang hilang dengan baik. Pohon keputusan yang dihasilkan oleh algoritma ini menawarkan visualisasi *rule* (aturan klasifikasi) yang sangat mudah diinterpretasikan oleh manusia, sehingga pemilik warung yang awam terhadap ilmu komputer sekalipun dapat memahami pola pembelian pelanggannya. Seperti penelitian berikut Menerapkan C4.5 pada lebih dari 13.000 data UMKM untuk mengklasifikasikan tingkat omzet (rendah, sedang, tinggi) berdasarkan atribut profil usaha, dengan tujuan mendukung sistem pendukung keputusan pembinaan UMKM [7].

Terdapat beberapa penelitian terdahulu yang relevan, penelitian [8] menganalisis akurat algoritma C4.5 dalam mengklasifikasi produk penjualan makanan terlaris. Pada Penelitian [9] mengimplementasikan Algoritma C4.5 untuk mengklasifikasikan penjualan makanan terlaris namun belum melibatkan parameter cuaca sebagai penentu keputusan dengan akurasi 84%. Disisi lain Penelitian [10] membandingkan algoritma Naive Bayes dan C4.5 dalam klasifikasi data penjualan Buku pada toko XYZ, di mana C4.5 terbukti memberikan hasil visualisasi yang lebih aplikatif bagi pemilik usaha. Selanjutnya, penelitian [11] memanfaatkan atribut Keramahan pelayanan (sangat ramah, ramah, biasa saja dan tidak ramah) untuk menentukan tingkat kepuasan layanan usaha kuliner tradisional menggunakan C4.5. Terakhir, menggunakan komparasi algoritma C4.5 dan K-Nearest Neighbor untuk mengklasifikasi penyakit kanker payudara termasuk jinak atau ganas dengan akurasi 96,71% [12].

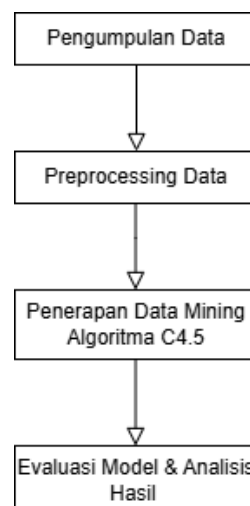
Oleh karena itu, kebaruan (*novelty*) dari penelitian ini dirumuskan untuk mengisi celah (*gap*) analisis tersebut. Kebaruan penelitian ini terletak pada pengintegrasian secara komprehensif antara variabel lingkungan eksternal dengan kebiasaan pelanggan ke dalam satu model prediksi. Variabel eksternal yang dilibatkan mencakup kondisi cuaca (Cerah/Hujan) dan tipe hari (*Weekday/Weekend*), yang digabungkan dengan jenis pesanan pelanggan (Bungkus/Makan di tempat). Pendekatan multidimensional ini diterapkan secara spesifik pada klasifikasi menu berbahan dasar *seafood* dan mie daerah, yang mana kombinasi tersebut masih sangat jarang dikaji pada penelitian-penelitian terdahulu.

Penelitian ini menggunakan parameter transaksi Juli–September 2025 untuk mengembangkan dan mengilustrasikan kemampuan Algoritma C4.5 dalam mengklasifikasikan menu favorit pelanggan UMKM Warung Mie Aceh *Seafood*. Penelitian ini bertujuan untuk menciptakan model pohon keputusan yang menghasilkan aturan kategorisasi (pengetahuan). *Rule* tersebut nantinya diharapkan dapat diimplementasikan oleh pihak UMKM sebagai landasan dalam sistem pendukung keputusan untuk merencanakan persediaan bahan pangan *seafood* harian yang lebih efektif [13], serta membantu manajemen dalam merumuskan strategi promosi yang lebih adaptif berdasarkan hari dan cuaca.

2. METODOLOGI PENELITIAN

2.1 Alur Penelitian

Tahapan penelitian ini dirancang secara sistematis untuk memastikan bahwa penyelesaian masalah klasifikasi menu favorit pada UMKM Warung Mie Aceh Seafood dapat dilakukan secara terstruktur dan terukur. Alur penelitian menggambarkan urutan implementasi metode mulai dari pengumpulan data hingga pengujian hasil. Berikut adalah visualisasi alur penelitian yang dilakukan:



Gambar 1. Tahapan Alur Penelitian

Gambar 1 merupakan tahapan penelitian yang dilakukan pada penelitian ini, dimulai dari pengumpulan data, *pre-processing* data, penerapan algoritma C4.5, evaluasi model & hasil klasifikasi.

a. Pengumpulan Data

Data yang digunakan merupakan data primer yang diperoleh dari historis transaksi penjualan selama tiga bulan (Juli–September 2025) sebanyak 586 entri. 9 atribut awal yang meliputi atribut Tanggal, Nama Menu, Jenis Menu, Harga, Jumlah Terjual, Total Penjualan, Jenis Pesanan, kondisi Cuaca, serta tipe Hari.

b. *Pre-processing* Data

Data yang dikumpulkan lanjut ke tahap *pre-processing* data agar difilter sesuai algoritma klasifikasi. Tahapan ini meliputi Data *Reduction* (pengurangan atribut) yaitu mengeliminasi atribut yang tidak relevan sebagai prediktor guna menghindari kondisi model yang terlalu menghafal data latih (*overfitting*) [14]. Transformasi Data yaitu mengubah data numeric menjadi kategorikal [15].

c. Penerapan Data Mining algoritma C4.5

Dataset bersih (*clean data*) diproses menggunakan perangkat lunak analitik *RapidMiner*. Data dibagi menggunakan metode pengujian untuk memisahkan data latih (*training data*) dan data uji (*testing data*). Setelah itu, algoritma pohon keputusan (*decision tree*) diterapkan pada data latih untuk membentuk model klasifikasi.

d. Evaluasi kinerja model dan Hasil Klasifikasi

Model yang dihasilkan dievaluasi untuk kinerja klasifikasi. Penelitian ini mengevaluasi matriks, akurasi, presisi, dan *recall* [8]. Selain itu, pendekatan validasi silang **k-fold** digunakan untuk memastikan generalisasi model dan tidak terjadi *overfitting* [16]. Hasil evaluasi kemudian dianalisis melalui interpretasi pohon keputusan dan aturan klasifikasi (*decision rules*) yang dihasilkan oleh algoritma C4.5. Analisis ini bertujuan untuk mengetahui atribut-atribut yang paling berpengaruh dalam menentukan menu favorit pelanggan serta memberikan dasar pengambilan keputusan bagi pemilik usaha kuliner berupa aturan (*rules*) keputusan.

2.2 Data Mining

Penambangan data mengekstrak informasi atau menemukan pola yang signifikan, inovatif, dan mungkin berharga dari kumpulan data besar [10]. UMKM Warung Mie Aceh Seafood akan menggunakan penemuan pengetahuan untuk menambang data transaksi hariannya, yang sebelumnya hanya berisi informasi penjualan. Penemuan Pengetahuan dalam Basis Data mengumpulkan dan menganalisis kumpulan data yang sangat besar untuk menemukan pola. Basis data melibatkan seleksi, pra-pemrosesan/pembersihan, transformasi, penambangan data, dan penilaian [17].

Langkah tersebut yang berawal dari seleksi data yang relevan, diteruskan dengan tahap pembersihan guna menyingkirkan nilai yang keliru, tidak utuh, ataupun tidak konsisten. Setelah itu, data yang telah bersih dikonversi ke dalam bentuk yang lebih tepat untuk dilakukan analisis. Pada fase selanjutnya, data diolah memakai algoritma spesifik demi membentuk model yang diperlukan. Akhirnya, performa model diperiksa untuk menjamin derajat akurasi serta relevansi pemanfaatannya dalam memecahkan problematika yang dikaji [18].

Klasifikasi adalah salah satu pekerjaan mendasar dalam penambangan data. Ini melibatkan pembangunan model yang menjelaskan dan memisahkan kelas data atau ide untuk memprediksi kelas objek dengan label kelas yang tidak diketahui [19][20]. Pendekatan ini dipilih karena selaras dengan tujuan penelitian untuk mengklasifikasikan kecenderungan pelanggan terhadap suatu menu.

2.3 Algoritma C4.5

Pemilihan algoritma C4.5 dalam penelitian ini didasarkan pada beberapa keunggulan strategis dibandingkan metode klasifikasi lain seperti Naive Bayes atau K-Nearest Neighbor (K-NN). Algoritma C4.5 memiliki kemampuan komputasi yang sangat prima dalam menangani atribut bertipe numerik dan kategorikal secara bersamaan [21], serta tangguh dalam menangani nilai data yang hilang atau tidak lengkap (*missing values*). Keunggulan utama lainnya adalah model keluaran C4.5 berupa pohon keputusan yang merepresentasikan struktur hierarki klasifikasi yang sangat sederhana dan mudah dipahami. Hasil akhirnya dapat diekstraksi menjadi aturan JIKA-MAKA (*IF-THEN rules*) dan Karakteristik ini sangat krusial, karena visualisasi dan aturan bisnis yang dihasilkan dapat langsung diinterpretasikan dan diimplementasikan oleh pemilik UMKM yang awam terhadap pemrograman dan statistika kompleks.

C4.5 adalah metode pohon keputusan berdasarkan ID3 [15]. ID3, yang diciptakan oleh J. Ross Quinlan, adalah metode induksi pohon keputusan seperti C.45 [22]. Pohon keputusan memecah sekumpulan data yang kompleks menjadi struktur hirarki klasifikasi yang lebih sederhana dan mudah dipahami. Dalam algoritma ini, pembentukan pohon dimulai dari simpul akar (*root node*), kemudian bercabang pada atribut penentu, dan berakhir pada daun (*leaf node*) yang merepresentasikan kelas target [23].

Untuk menentukan atribut mana yang akan menjadi akar atau cabang, algoritma C4.5 menggunakan dua konsep matematis utama, yaitu Entropy dan Information Gain [24]. Penomoran rumus yang digunakan dalam perhitungan ini yaitu :

a. Menghitung Entropy :

Entropy untuk mengukur tingkat ketidakpastian atau keheterogenan dari kumpulan data.

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

Keterangan:

- S : Himpunan Kasus
- N : Jumlah Partisi S
- Pi : Proporsi dari Si Terhadap S

b. Menghitung Information Gain :

Setelah Entropy dihitung, algoritma akan mencari nilai Information Gain tertinggi dari setiap atribut untuk dijadikan simpul keputusan.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} Entropy(S_i) \quad (2)$$

Keterangan :

- S : Himpunan Kasus
- A : Atribut
- N : Jumlah Partisi A
- |Si| : Jumlah Kasus pada Partisi ke-i
- |S| : Jumlah Kasus Pada S

2.4 Evaluasi Model dan Hasil Klasifikasi

Hasil klasifikasi merupakan representasi pengetahuan yang diperoleh setelah algoritma C4.5 memproses kumpulan training data. Atribut yang digunakan untuk klasifikasi menu favorit yaitu 3 atribut predaktor yang meliputi, Jenis Menu, Cuaca dan Hari. Pada penelitian ini, atribut numerik jumlah terjual ditransformasikan kedalam kelas target (Label_Target) yaitu: "Sangat Favorit", "Favorit", dan "Tidak Favorit".

Berikut ini implementasi parameter pengujian RapidMiner Untuk menjamin keandalan hasil yang terbebas dari bias pembagian data, pengujian menggunakan operator Validasi Silang (*Cross-Validation*). Pengujian menggunakan metode K-Fold dengan nilai parameter k = 10 karena metode ini terbukti mampu memberikan estimasi performa yang lebih stabil, objektif, dan representatif dalam menguji keandalan algoritma C4.5 [2]. Parameter yang dikonfigurasi pada perangkat lunak dijabarkan sebagai berikut:

a. Parameter *Cross-Validation*

Pengujian menggunakan metode **K-Fold** dengan nilai parameter k = 10, serta menggunakan tipe pengambilan sampel otomatis (*automatic sampling type*). Konfigurasi ini secara otomatis membagi dataset acak menjadi 10 bagian pemerataan (9 bagian sebagai data latih dan 1 bagian sebagai data uji yang diputar secara iteratif).

b. Konfigurasi *Sub-Proses Training*

Pada blok pelatihan, digunakan operator Decision Tree (C4.5). Parameter *criterion* diatur pada perhitungan information gain (perolehan informasi) untuk memastikan simpul pohon dihitung berdasarkan rumus matematis C4.5.

c. Konfigurasi *Sub-Proses Testing*

Pada sisi pengujian, digunakan operator *Apply Model* yang berfungsi menerima model terlatih dan menerapkannya pada data uji untuk menghasilkan prediksi kelas target.

Hasil akhir dari klasifikasi ini tidak hanya berupa struktur visual pohon keputusan, melainkan dapat diekstraksi menjadi seperangkat aturan "JIKA-MAKA" (*IF-THEN rules*) [25]. Aturan klasifikasi inilah yang akan dievaluasi kinerjanya menggunakan *Confusion Matrix* seperti penelitian ini [26] untuk memastikan model dapat memprediksi label kelas dengan presisi dan akurasi yang tinggi saat disajikan data kondisi cuaca (Cerah/Hujan) atau hari (*Weekday/Weekend*) di masa mendatang.

3. HASIL DAN PEMBAHASAN

3.1 Deskripsi Dataset Penelitian

Data yang digunakan dalam penelitian ini merupakan data historis transaksi harian yang didapatkan langsung dari *Point of Sales* (POS) atau pencatatan penjualan UMKM Warung Mie Aceh Seafood. Observasi data ditarik dari periode tiga bulan berturut-turut, yakni dari awal bulan Juli hingga akhir September 2025, dengan total data yang di peroleh sebanyak 586 entri. Atribut yang mencakup dalam dataset ini memiliki perekaman variabel yang cukup detail mengenai operasional warung, yang mencakup 9 atribut awal, yaitu: Tanggal, Nama Menu, Jenis Menu, Harga, Jumlah Terjual, Total Penjualan, Jenis Pesanan (Bungkus atau Makan di tempat), Cuaca (Cerah atau Hujan), serta Hari (*Weekday* atau *Weekend*). Seluruh data tersebut dimanfaatkan sebagai dasar analisis dalam membangun model klasifikasi menu favorit pelanggan dengan menerapkan metode C4.5. dimana penentuan variable target dilakukan sesuai dengan tujuan penelitian. Setelah data terkumpul dilakukan tahapan *pre-processing* data menggunakan *Tolls Rapidminer*

3.2 Tahap Pre-processing Data

Data mentah yang diperoleh dari UMKM tidak dapat secara langsung dimasukkan ke dalam model perhitungan C4.5 karena masih memiliki redundansi informasi dan format atribut numerik target yang belum sesuai. Oleh karena itu, dalam penelitian ini tahap *pre-processing* data dilakukan dengan dua tahap utama.

a. *Data Reduction* (Pengurangan Atribut)

Pada penelitian ini, pengurangan atribut Tanggal, Harga, Total Penjualan, dan Jenis Pesanan dieliminasi. Penghapusan ini bertujuan agar model tidak mengalami *overfitting* dan dapat berfokus murni pada korelasi antara variabel lingkungan (Cuaca, Hari) dan karakteristik produk (Jenis Menu) terhadap minat beli.

b. *Data Transformation*

Transformasi dilakukan dengan pembuatan atribut target baru bernama "Label_Target" yang ditransformasikan dari atribut numerik "Jumlah Terjual". Pembagian kategori kelas (label) didasarkan pada perhitungan ambang batas performa warung dengan parameter sebagai berikut:

1. Tidak Favorit: Jika Jumlah Terjual ≤ 11 porsi.
2. Favorit: Jika Jumlah Terjual berkisar 12 hingga 18 porsi.
3. Sangat Favorit: Jika Jumlah Terjual ≥ 19 porsi.

Berdasarkan dari hasil pembagian kategori kelas tersebut, maka diperoleh kelas target sebagai berikut :

Tabel 1. Kelas target

No	Label Target	Jumlah Data	Persentase %
1.	Sangat Favorit	173	29,52 %
2.	Favorit	293	50 %
3.	Tidak Favorit	120	20,48 %
Total		586	100 %

Tabel 1 Berdasarkan hasil distribusi data, terlihat bahwa kategori "Favorit" mendominasi dataset dengan proporsi sebesar 50% (293 data). Sementara itu, kategori ("Sangat Favorit") dan ("Tidak Favorit") memiliki proporsi yang lebih kecil namun masih dalam rentang yang wajar, yakni berturut-turut sebesar 29,52% (173 data) dan 20,48% (120 data). Kondisi ini menunjukkan bahwa dataset memiliki distribusi kelas yang tergolong cukup seimbang meskipun terdapat dominasi pada kelas Favorit. Dengan demikian, model klasifikasi C4.5 dapat langsung dibangun secara optimal tanpa memerlukan teknik penanganan khusus terhadap ketidakseimbangan data (*data imbalance*).

Berikut ini data yang dihasilkan setelah melewati tahap *pre-processing*, menghapus atribut tanggal, harga dan total penjualan yang dianggap tidak relevan dalam proses klasifikasi. Dataset ini terdiri dari 9 atribut, termasuk atribut target, yaitu Status Favorit. Penelitian ini menggunakan atribut mencakup jumlah terjual, jenis menu, jenis pesanan, cuaca, hari dan label_target.

Tabel 2. Hasil Preprocessing Data

Jenis Menu	Jenis Pesanan	Cuaca	Hari	Label Target
Mie Instan	Bungkus	Cerah	Weekday	Tidak Favorit
Mie Aceh	Bungkus	Cerah	Weekday	Favorit
Jajanan	Bungkus	Cerah	Weekday	Sangat Favorit
Jajanan	Makan di tempat	Cerah	Weekday	Sangat Favorit
Mie Aceh	Makan di tempat	Cerah	Weekday	Favorit
Mie Aceh Seafood	Bungkus	Hujan	Weekend	Favorit
.....				
Minuman	Bungkus	Hujan	Weekend	Tidak Favorit
Mie Aceh	Makan di tempat	Hujan	Weekday	Favorit

Berdasarkan Tabel 2 di atas, dataset kini memiliki struktur yang lebih ringkas dan siap diolah oleh sistem. Label kelas target pada data tersebut telah berhasil dikelompokkan menjadi tiga kategori prediksi, yaitu 'Sangat Favorit', 'Favorit', dan 'Tidak Favorit'. Ratusan rekam data bersih ini selanjutnya akan digunakan sebagai basis data latih (*training data*) pada proses klasifikasi algoritma C4.5 untuk menghitung nilai *Entropy* dan *Gain*, yang bermuara pada pembentukan struktur pohon keputusan yang sesuai.

Sebuah penelitian terdahulu menegaskan bahwa keberhasilan model klasifikasi sangat dipengaruhi oleh kualitas *pre-processing* data, khususnya pada metode klasifikasi berbasis pohon keputusan seperti decision tree C4.5[27].

Berdasarkan kondisi tersebut, penelitian ini menjadi relevan untuk dilakukan untuk memastikan bahwa hasil *pre-processing* data memenuhi kebutuhan teknis untuk pembentukan model klasifikasi. Tujuan utama penelitian ini adalah menggunakan data yang telah disusun dengan benar untuk membangun model klasifikasi menggunakan algoritma Decision Tree C4.5 sesuai dengan karakteristik data dan tujuan analisis yang telah ditetapkan.

3.3 Perhitungan Entropy dan Gain

Sebelum menentukan atribut terbaik dalam pembentukan pohon keputusan, maka dari itu terlebih dahulu menghitung nilai *Entropy* total dataset. *Entropy* digunakan untuk mengukur tingkat ketidakpastian distribusi kelas pada data awal. lalu kemudian dilanjutkan dengan pencarian nilai *Information Gain* pada masing-masing atribut prediktor yaitu Jenis Menu, Cuaca, Hari.

Berdasarkan tabel 2, dengan total data yang diperoleh sebanyak 586 data transaksi.

- a. Sangat Favorit = 173 data

- b. Favorit = 293 data
c. Tidak Favorit = 120 data

Berikut ini Rumus perhitungan entropy total:

$$Entropy(S) = - \sum_{i=0}^n p_i \log_2 p_i \quad (3)$$

Subtitusikan probabilitas masing-masing kelas:

$$Entropy(S) = - \left(\frac{173}{586} \log_2 \frac{173}{586} \right) - \left(\frac{293}{586} \log_2 \frac{293}{586} \right) - \left(\frac{120}{586} \log_2 \frac{120}{586} \right)$$

Perhitungan :

$$\begin{aligned} &= -(0,2952 \log_2 0,2952) - (0,5 \log_2 0,5) - (0,2048 \log_2 0,2048) \\ &= -(0,2952 \times -1,760) - (0,5 \times -1) - (0,2048 \times -2,2877) \\ &= -(0,5196 + 0,5 + 0,4685) \\ &Entropy(S) = 1,4881 \end{aligned}$$

Maka didapatkan Nilai Entropy total sebesar 1,4881. Berdasarkan metode perhitungan entropy dan gain yang telah dijelaskan, selanjutnya dilakukan perhitungan terhadap dataset penelitian untuk menentukan atribut terbaik dalam pembentukan pohon keputusan.

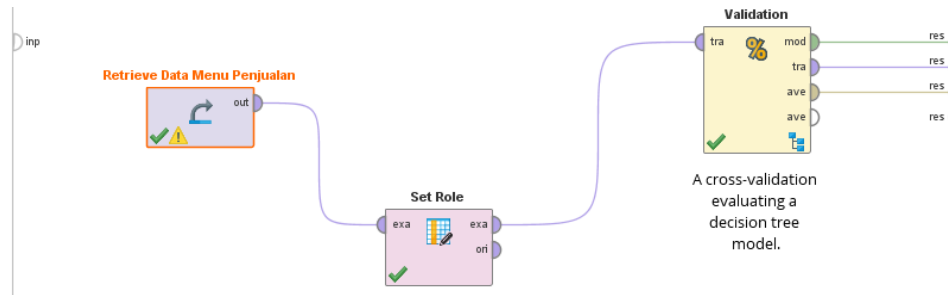
Tabel 3. Hasil Perhitungan Algoritma C4.5

Atribut	Value	Jumlah Kasus	Sangat Favorit	Favorit	Tidak Favorit	Entropy	Gain
Total		586	173	293	120	1,488133453	
Jenis Menu							0,374957707
	Mie Aceh	145	47	98	0	0,908824671	
	Mie Aceh	71	6	48	17	1,176862002	
	Seafood	103	28	26	49	1,522052441	
	Mie Instan	25	0	25	0	0	
	Nasi Goreng	53	53	0	0	0	
	Jajanan	189	39	96	54	1,482607442	
Jenis pesanan							0,016004967
	Bungkus	309	73	174	62	1,423276839	
	Makan di tempat	277	100	119	58	1,526623644	
Cuaca							0,044572561
	Cerah	415	116	237	62	1,385357607	
	Hujan	171	57	56	58	1,584814479	
Hari							0,160938791
	Weekday	420	66	248	106	1,369649649	
	Weekend	166	107	45	14	1,219778433	

Berdasarkan hasil komputasi dari dataset UMKM Warung Mie Aceh Seafood, atribut "Jenis Menu" menghasilkan nilai *Gain* tertinggi dibandingkan atribut lainnya, sehingga sistem memposisikan "Jenis Menu" sebagai akar atau *Root Node*. Cabang-cabang dari *Root Node* ini kemudian dipecah lagi menggunakan atribut "Hari" dan "Cuaca" yang bertindak sebagai *Internal Node*, hingga seluruh cabang bermuara pada kesimpulan akhir di daun (*LeafNode*) yang mendeskripsikan *Label Target* (Sangat Favorit, Favorit, Tidak Favorit).

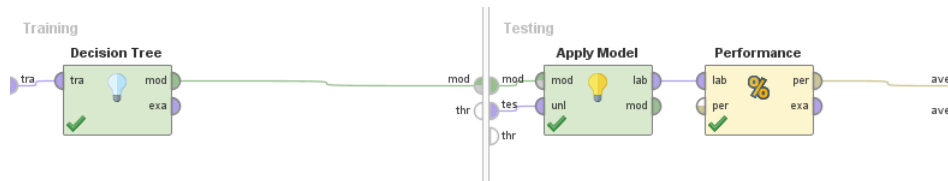
3.4 Pembentukan Pohon Keputusan

Dataset yang telah melalui *Pre-processing* dimasukkan kedalam *software/ tools RapidMiner* untuk membentuk pohon keputusan (*decision tree*), menghitung nilai *akurasi*, *presisi* dan *recall* dengan algoritma C4.5. berikut ini desain model yang digunakan untuk klasifikasi.



Gambar 2. Proses Utama di RapidMiner

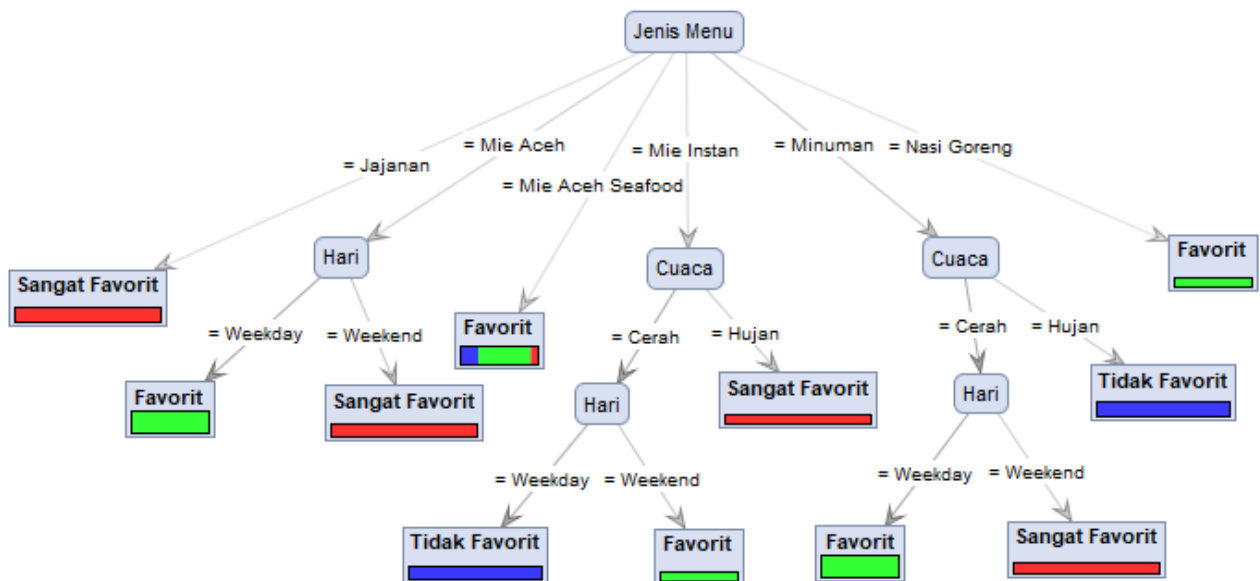
Penelitian ini menggunakan *Cross Validation* yang digunakan data secara acak untuk menentukan data *testing* dan data *training* menggunakan **k-Fold 10** dan *sampling type otomatic*.



Gambar 3. Konfigurasi Sub-Proses Training dan Testing pada Cross Validation

Alur validasi silang (*cross validation*) pada Gambar 3 memastikan bahwa model dilatih secara rotasional sehingga kemampuannya dalam memprediksi data uji menjadi objektif dan terbebas dari *overfitting*. Hasil dari pengolahan model tersebut menghasilkan visualisasi struktur hierarki pohon keputusan yang disajikan pada Gambar 4.

Tabel 3 menunjukkan bahwa Jenis Menu memiliki nilai gain terbesar, menjadikannya simpul akar pohon keputusan. Hal ini menunjukkan bahwa tipe menu mendominasi minat konsumen, yang kemudian didukung oleh kondisi lingkungan spesial-temporal seperti Hari dan Cuaca pada *node-node* percabangan selanjutnya.



Gambar 4. Pohon Keputusan Klasifikasi Menu Favorit

Gambar 4 belum sepenuhnya merepresentasikan wawasan yang siap pakai bagi UMKM. Oleh karena itu, struktur visualisasi pohon keputusan tersebut diekstraksi ke dalam himpunan pengetahuan (*knowledge representation*) berupa 11 aturan keputusan (*IF-THEN rules*) seperti yang dirangkum pada tabel 4.

Tabel 4. Hasil Rules Aturan Keputusan

No	Rules	Keputusan
1.	IF Jenis Menu = Jajanan	Sangat Favorit
2.	IF Jenis Menu = Mie Aceh AND Hari = Weekday	Favorit
3.	IF Jenis Menu = Mie Aceh AND Hari = Weekend	Sangat Favorit
4.	IF Jenis Menu = Mie Aceh Seafood	Favorit

5.	IF Jenis Menu = Mie Instan AND Cuaca = Cerah AND Hari Weekday	Tidak Favorit
6.	IF Jenis Menu = Mie Instan AND Cuaca = Cerah AND Hari Weekend	Favorit
7.	IF Jenis Menu = Mie Instan AND Cuaca = Hujan	Sangat Favorit
8.	IF Jenis Menu = Minuman AND Cuaca = Cerah AND Hari Weekday	Favorit
9.	IF Jenis Menu = Minuman AND Cuaca = Cerah AND Hari Weekend	Sangat Favorit
10.	IF Jenis Menu = Minuman AND Cuaca = Hujan AND	Tidak Favorit
11.	IF Jenis Menu = Nasi Goreng	Favorit

Tabel 4 menyajikan pola perilaku konsumen yang dapat diterjemahkan ke dalam strategi operasional bisnis riil. Sebagai contoh, Aturan 7 dan 10 memperlihatkan adanya korelasi substitusi yang kuat: ketika "Cuaca = Hujan", "Mie Instan" berubah menjadi "Sangat Favorit", sedangkan "Minuman" justru "Tidak Favorit" Aturan ini membuktikan hipotesis bahwa variabel cuaca sangat memengaruhi selera pelanggan terhadap menu tertentu.

Aturan keputusan tersebut dapat digunakan sebagai dasar menentukan rekomendasi jenis menu yang paling diminati pelanggan.

3.4 Hasil Penelitian

Untuk memastikan bahwa aturan (*rules*) yang didapatkan pada tahap sebelumnya cukup reliabel untuk diimplementasikan, performa model dievaluasi dengan teknik pengujian *Confusion Matrix*. Evaluasi ini memetakan nilai prediksi sistem melawan nilai kelas aktual di lapangan. Visualisasi matriks performa sistem disajikan pada **Gambar 2**.

Untuk melihat performa klasifikasi lebih rinci, berikut hasil pengujian menggunakan *software RapidMiner* yang disajikan dalam bentuk *confusion Matrix*.

Tabel 5. Hasil Pengujian *Confusion Matrix*

Prediksi / Aktual	Tidak Favorit	Favorit	Sangat Favorit	Class Precision
Tidak Favorit	103	0	0	100.00 %
Favorit	17	293	6	92.72 %
Sangat Favorit	0	0	167	100.00 %
Class Recall	85.83 %	100.00 %	95.53 %	

Dari hasil kalkulasi *Confusion Matrix* pada **Tabel 5**, model algoritma C4.5 mencapai performa prediksi sangat prima, Akurasi sejumlah 96,08% (+/- 2,52%). Secara rincian kelas, sistem mampu mengidentifikasi data kelas "Tidak Favorit" sebanyak 103 kasus aktual secara tepat tanpa meleset (*Class Precision* 100,00%). Prediksi untuk menu "Sangat Favorit" juga sangat presisi di mana 167 prediksi seluruhnya tepat sasaran (*Class Precision* 100,00%). Sementara untuk klasifikasi "Favorit", dari keseluruhan prediksi terdapat tingkat *recall* sebesar 100,00% yang berarti sistem tidak pernah gagal mengenali kondisi saat suatu menu seharusnya berstatus favorit.

Tingginya tingkat akurasi pada model ini membuktikan keunggulannya jika dikomparasikan dengan studi-studi terdahulu. Pada penelitian sejenis sebelumnya yang menggunakan algoritma C4.5 untuk klasifikasi kelasan makanan namun tanpa mengintegrasikan variabel cuaca, tingkat akurasi yang dihasilkan hanya berkisar di angka 84% Peningkatan akurasi menjadi 96,08% pada penelitian ini membuktikan secara ilmiah bahwa integrasi parameter lingkungan eksternal ("Cuaca" dan "Hari") sangat krusial dan secara signifikan menajamkan ketepatan model prediktif. Tingkat akurasi yang mendekati sempurna ini membuktikan bahwa kombinasi parameter "Jenis Menu", "Cuaca", dan "Hari" adalah prediktor yang valid terhadap frekuensi pemesanan pelanggan.

Hasil penelitian ini memberikan luaran analisis (*insight*) bisnis yang sangat krusial bagi UMKM Warung Mie Aceh Seafood:

a. Perilaku Konsumsi Jajanan

Kategori "Jajanan" (seperti bakso bakar dan bakso kuah kari) mutlak selalu berada di tingkat permintaan tertinggi ("Sangat Favorit"), tidak terpengaruh oleh anomali cuaca maupun hari libur. Hal ini mengindikasikan bahwa jajanan merupakan menu pendamping konstan bagi pelanggan.

b. Sensitivitas Cuaca terhadap Mie Instan dan Minuman

Terdapat pola substitusi yang berlawanan antara Mie Instan dan Minuman. Saat cuaca hujan, pesanan Mie Instan akan meledak menjadi "Sangat Favorit", namun di saat bersamaan pesanan produk Minuman (seperti es rasa-rasa) akan anjlok menjadi "Tidak Favorit".

c. Pola Pemesanan di Akhir Pekan

Menu andalan "Mie Aceh" beserta produk komplementernya mengalami eskalasi penjualan dari sekadar "Favorit" pada hari kerja menjadi "Sangat Favorit" pada hari libur (*Weekend*). Fenomena serupa juga terjadi pada produk "Minuman" jika akhir pekan tersebut cerah.

Dengan tergalinya pola pengetahuan tersebut, manajemen UMKM sangat mempercayai model keputusan ini untuk mencegah penumpukan. maka dari itu tindakan optimal yang dapat diimplementasikan adalah mengurangi produksi

es batu dan persiapan minuman saset secara drastis pada saat prakiraan cuaca menunjukkan hujan lebat, dan segera mendiversifikasi modal tersebut untuk menambah pasokan persediaan Mie Instan. Selain itu, manajemen persediaan (*inventory*) bahan baku Mie Aceh dan produk Jajanan harus selalu diprioritaskan pada kuantitas maksimum setiap menjelang pergeseran waktu hari kerja menuju akhir pekan (*Weekend*) untuk mencegah potensi hilangnya pendapatan (*lost sales*) akibat kehabisan stok. Implementasi dari penemuan pola ini sejalan dengan konsep rekayasa menu, di mana hasil klasifikasi menu kuliner dapat secara langsung digunakan sebagai rekomendasi strategis bagi pihak manajemen UMKM dalam merancang strategi pemasaran, penyesuaian harga, serta operasional harian guna meningkatkan keuntungan bisnis [28].

Untuk menjaga objektivitas dan kualitas ilmiah penelitian, terdapat beberapa kelemahan pada model yang dikembangkan ini yang perlu dipertimbangkan:

a. Keterbatasan Variabel Demografi dan Waktu Spesifik

Atribut yang digunakan untuk memprediksi baru sebatas faktor Cuaca dan Hari Model ini lemah dalam mengidentifikasi pola kebiasaan personal karena belum melibatkan variabel demografi pelanggan (usia/pekerjaan) atau rentang jam kunjungan spesifik (pagi, siang, malam), yang bisa jadi memengaruhi variasi porsi pesanan menu

b. Periode Latih Data yang Singkat

Model ini dilatih menggunakan lingkup waktu yang terbatas, yakni data operasional selama tiga bulan Durasi ini berpotensi belum merekam anomali tren pemesanan musiman, seperti bulan puasa, masa libur panjang sekolah, atau libur akhir tahun.

c. Tidak Adanya Uji Komparasi Internal

Penelitian ini hanya menerapkan algoritma tunggal (C4.5) tanpa melakukan uji banding performa secara internal dengan algoritma klasifikasi sejenis (seperti Naive Bayes atau Random Forest) terhadap dataset yang sama.

Oleh karena itu, akurasi 96,08% belum dapat diklaim secara absolut sebagai satu-satunya algoritma terbaik untuk kasus spesifik pada warung ini.

4. KESIMPULAN

Penelitian ini membuktikan bahwa algoritma C4.5 sangat efektif diimplementasikan untuk mengklasifikasikan menu favorit pelanggan pada UMKM Warung Mie Aceh Seafood melalui pengujian terhadap 586 rekam data transaksi menggunakan metode 10-fold cross validation, model prediksi ini terbukti sangat andal dengan pencapaian akurasi sebesar 96,08%, serta class precision sempurna (100%) pada kategori "Sangat Favorit" dan "Tidak Favorit" kontribusi utama dari penelitian ini adalah terekstraksinya 11 aturan keputusan (*rules*) yang memberikan wawasan strategis praktis bagi UMKM. Pengetahuan baru (*knowledge*) yang dihasilkan dari aturan klasifikasi tersebut memberikan wawasan strategis bagi pihak manajemen warung. Diketahui bahwa menu 'Jajanan' memiliki tingkat permintaan tinggi yang konstan tanpa terpengaruh oleh faktor cuaca maupun hari libur, sedangkan pemesanan 'Mie Instan' dan 'Minuman' sangat fluktuatif serta berbanding terbalik terhadap kondisi cuaca hujan. Keterbatasan dari penelitian ini adalah atribut prediktor yang digunakan masih terbatas pada faktor cuaca dan hari, serta lingkup data yang hanya mencakup waktu tiga bulan operasional. Untuk pengembangan penelitian selanjutnya, disarankan agar peneliti menambahkan variabel atribut lain yang lebih kompleks, seperti data demografi pelanggan, jam kunjungan (pagi/siang/malam), serta mengadakan komparasi dengan algoritma klasifikasi lainnya ialah Random Forest atau Naive Bayes untuk melihat perbandingan akurasi yang lebih komprehensif.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada pihak UMKM Warung Mie Aceh Seafood yang telah bersedia memberikan data penjualan harian sebagai bahan penelitian. Penulis juga menyampaikan apresiasi kepada dosen pembimbing dan pihak program studi yang telah memberikan arahan, masukan dan saran serta dukungan selama proses penyusunan artikel ini. Selain itu, penulis berterima kasih kepada semua pihak yang telah turut membantu dalam pengumpulan data, proses pengolahan data dalam penyusunan artikel ini sehingga dapat penulis selesaikan dengan baik.

REFERENCES

- [1] M. A. N. Syifa, N. Marlyana, and W. Spalanzani, "Penerapan Metode Always Better Control (ABC) Terhadap Pengendalian Persediaan Bahan Baku Perishable Olahan Hasil Laut Pada UD.XYZ," *J. Jar. SainTek*, vol. 7, no. 1, pp. 7–14, 2025.
- [2] N. P. Rahmawati, I. Budiman, M. I. Mazdadi, A. Farmadi, and Dkk, "Comparison Between K-Fold Cross Validation And Percentage Split In Decision Tree Algorithms For Anemia Classification," *Indones. J. Electron. Electromed. Eng. Med. Informatics*, vol. 8, no. 1, pp. 105–114, 2026.
- [3] N. Groene and S. Zakharov, "Introduction of AI - based sales forecasting : how to drive digital transformation in food and beverage outlets," *Discov. Artif. Intell.*, 2024, doi: 10.1007/s44163-023-00097-x.
- [4] F. A. Karubun, D. B. Kluman, R. Febriansyah, P. E. Pali, and Dkk, "Prediction of Unselling Menu Items for Bundling and Promotion Strategies in Cafés Using Decision Tree Algorithm," *J. Ilm. Sist. Inf.*, vol. 5, no. 1, pp. 96–106, 2026.
- [5] E. A. Yestina, B. H. Prakoso, E. Selviyanti, and G. E. J. Eko, "Implementasi Algoritma C4 . 5 pada Analisis Faktor Risiko Penyakit Jantung Koroner," *J. Ris. Komput.*, vol. 12, no. 5, pp. 684–694, 2025, doi: 10.30865/jurikom.v12i5.8852.
- [6] N. Khasanah, D. U. E. Saputri, T. Hidayat, F. Aziz, and Dkk, "Prediksi Kelulusan Mahasiswa Menggunakan Algoritma C4 . 5

- dengan RapidMiner : Studi Kasus Data Akademik Perguruan Tinggi XYZ,” *Indones. J. Comput. Sci.*, vol. 4, no. 2, pp. 100–107, 2025.
- [7] Y. F. Saputra and S. Sugiarto, “Algoritma C4.5 sebagai Penerapan Decision Tree-Based Classification Model untuk Mengklasifikasikan Tingkat Omzet UMKM Berdasarkan Profil Bisnis,” *J. Ekon.*, vol. 13, no. 2, pp. 188–197, 2024.
 - [8] F. Rahmadayanti, A. Asminah, and D. T. Cahaya, “Penerapan Algoritma C4.5 Pada Produk Penjualan Makanan Ringan,” *J. Teknol. Inf. Mura*, vol. 16, no. 1, pp. 62–69, 2024, doi: 10.32767/jti.v16i1.2294.
 - [9] F. G. Falentina, G. Y. Y. Wabdaron, D. Andiyani, and Dkk, “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. 385–395, 2025.
 - [10] E. Rianty, K. Budi, and Effiyaldi, “Perbandingan Kinerja Algoritma C4 . 5 dan Naive Bayes Dalam Klasifikasi Data Penjualan Buku PT . XYZ,” *J. Ris. Komput.*, vol. 12, no. 6, pp. 793–804, 2025, doi: 10.30865/jurikom.v12i6.9345.
 - [11] L. N. Ningsih, R. Septiani, A. Pramadjaya, and Dkk, “Implementing the C4 . 5 Algorithm for Customer Satisfaction Classification,” *Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. October, pp. 1470–1480, 2025.
 - [12] I. Muhammad, M. Donny, H. Wahyu, K. N. Tedi, and Dkk, “Analysis of K-Nearest Neighbor (KNN), Naive Bayes ands Decision Tree C4 . 5 Algorithm With Classification Method In Breast Cancer Using RapidMiner,” *J. Appl. Intell. Syst.*, vol. 9, no. 2, pp. 260–270, 2024.
 - [13] P. Li, F. Xiong, X. Huang, and X. Wen, “Construction and optimization of vending machine decision support system based on improved C4 . 5 decision tree,” *Heliyon*, vol. 10, no. 3, p. e25024, 2024, doi: 10.1016/j.heliyon.2024.e25024.
 - [14] A. K. B. Ginting, M. S. Lydia, and E. M. Zamzami, “Reduksi Atribut Menggunakan Chi Square untuk,” *J. Edukasi dan Penelit. Inform.*, vol. 9, no. 1, pp. 44–49, 2023.
 - [15] M. Fadilah and L. Tanti, “Klasifikasi Jenis Usaha Mikro Kecil Menengah (UMKM) Berdasarkan Skala Usaha Menggunakan Algoritma C4.5,” *SENADIMU*, vol. 2, no. 1, pp. 814–827, 2025.
 - [16] N. D. Saputri, K. Khalid, and D. Rolliawati, “Komparasi Penerapan Metode Bagging dan Adaboost pada Algoritma C4 . 5 untuk Prediksi Penyakit Stroke,” *J. Sistm Inf.*, vol. 11, no. September, pp. 567–577, 2022.
 - [17] F. Zafira, B. Irawan, and A. Bahtiar, “Penerapan Data Mining Untuk Estimasi Stok Barang Dengan Metode K-Means Clustering,” *J. Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 156–161, 2024.
 - [18] M. B. Pratama and F. I. Sanjaya, “Prediksi omset penjualan toko mbak ning pasar cinderamata menggunakan long short term memory,” *J. Sist. Inf. dan Sist. Komput.*, vol. 10, no. 1, pp. 160–167, 2025.
 - [19] A. H. Nasrullah, “Implementasi Algoritma Decision Tree Untuk Klasifikasi Produk laris,” *J. Ilm. Ilmu Komput.*, vol. 7, no. 2, pp. 45–51, 2021.
 - [20] J. Han, J. Pei, and H. Tong, *Data Mining Concepts And Techniques*. 2023.
 - [21] A. A. Alharbi, “Classification Performance Analysis of Decision Tree-Based Algorithms with Noisy Class Variable,” *Hindawi*, vol. 2024, pp. 1–10, 2024, doi: 10.1155/2024/6671395.
 - [22] T. Widodo, D. Setiawan, and A. Syahputri, “Data Mining Menentukan Minat Konsumen Memilih Sepeda Motor Idaman Dengan Algoritma C4.5,” *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 1, no. 6, pp. 826–836, 2022, doi: 10.53513/jursi.v1i6.7262.
 - [23] P. B. N. Setio, D. R. Saputro, and B. Winarto, “Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C4.5,” *Pros. Semin. Nas. Mat.*, vol. 3, pp. 64–71, 2020.
 - [24] S. Suhartini, V. N. Permatasari, and I. Santoso, “Klasifikasi Penjualan Tempe Dengan Data Mining Menggunakan Algoritma C4.5,” *J. IKRAITH-INFORMATIKA*, vol. 7, no. 2, pp. 124–132, 2023.
 - [25] J. Han, M. Kamber, and J. Pei, “Data Mining : Concepts and Techniques Third Edition,” 2023.
 - [26] A. F. Azmi and A. Voutama, “Prediksi Churn Nasabah Bank Menggunakan Klasifikasi Random Forest dan Decision Tree Dengan Evaluasi Confusion Matrix,” *J. Ilm. Komput. dan Inform.*, vol. 13, no. 1, 2024.
 - [27] T. Gori, A. Sunyoto, and H. Al Fatta, “Preprocessing Data Dan Klasifikasi Untuk Prediksi Kinerja Akademik Siswa,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 1, pp. 215–224, 2024, doi: 10.25126/jtiik.20241118074.
 - [28] K. Dhanushkodi, A. Bala, N. Kodipyaka, and V. Shreyas, “Customer Behaviour Analysis and Predictive Modelling in Supermarket Retail: A Comprehensive Data Mining Approach,” *IEEE Access*, vol. PP, p. 1, 2024, doi: 10.1109/ACCESS.2024.3407151.