

KLASIFIKASI PELANGGAN SMARTPHONE SAMSUNG MENGUNAKAN DESISSION TREE PADA RAYA SELULER SOLO

Puspita Auliana Putri, Afu Ichsan Pradana, Bondan Wahyu Pamekas

Teknik Informatika, Universitas Duta Bangsa Surakarta

Jl. Bhayangkara, No.55, Tipes, Serengan, Surakarta

Sitiitaa76@gmail.com

ABSTRAK

Perkembangan penjualan smartphone yang terus meningkat menghasilkan data pelanggan dalam jumlah besar yang berpotensi dimanfaatkan untuk mendukung pengambilan keputusan bisnis. Namun, Raya Seluler Solo masih mengalami kesulitan dalam mengidentifikasi karakteristik pelanggan smartphone Samsung berdasarkan data transaksi yang tersedia. Kondisi tersebut menyebabkan strategi pemasaran dan pelayanan pelanggan belum berjalan secara optimal dan tepat sasaran. Penelitian ini bertujuan untuk mengklasifikasikan pelanggan smartphone Samsung pada Raya Seluler Solo menggunakan algoritma Decision Tree guna mendukung pengambilan keputusan berbasis data. Metode yang digunakan adalah data mining dengan algoritma Decision Tree C4.5 melalui tahapan selection, preprocessing, transformation, data mining, dan evaluation. Data penelitian diperoleh dari data pelanggan dan transaksi penjualan smartphone Samsung dengan atribut usia, jenis kelamin, pekerjaan, penghasilan, tipe smartphone, dan frekuensi pembelian. Hasil penelitian menunjukkan bahwa algoritma Decision Tree mampu mengklasifikasikan pelanggan ke dalam kategori pelanggan loyal, pelanggan potensial, dan pelanggan tidak loyal. Model yang dihasilkan memberikan aturan klasifikasi yang mudah dipahami sehingga dapat membantu pihak toko dalam mengenali karakteristik pelanggan dan menyusun strategi pemasaran yang lebih efektif. Dengan demikian, metode Decision Tree dapat digunakan sebagai solusi untuk mendukung pengelolaan data pelanggan pada Raya Seluler Solo.

Kata kunci : *Data Mining, Decision Tree, C4.5, Klasifikasi Pelanggan, Smartphone Samsung.*

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah mendorong peningkatan penggunaan smartphone sebagai perangkat yang mendukung berbagai aktivitas masyarakat, seperti komunikasi, pendidikan, pekerjaan, hiburan, dan transaksi digital. Tingginya kebutuhan terhadap smartphone menyebabkan persaingan antarprodusen semakin meningkat. Samsung merupakan salah satu merek smartphone yang memiliki pangsa pasar besar di Indonesia dengan berbagai varian produk yang dapat menjangkau beragam segmen konsumen. Kondisi tersebut menjadikan data pelanggan sebagai aset penting yang dapat dimanfaatkan untuk memahami perilaku pembelian dan mendukung penyusunan strategi pemasaran yang lebih efektif.

Raya Seluler Solo merupakan salah satu toko yang menjual berbagai tipe smartphone Samsung dengan jumlah transaksi yang terus meningkat. Seiring bertambahnya transaksi, data pelanggan yang tersimpan juga semakin banyak. Namun, data tersebut masih digunakan sebagai arsip transaksi dan belum dimanfaatkan secara optimal sebagai sumber informasi untuk mendukung pengambilan keputusan. Akibatnya, pihak toko mengalami kesulitan dalam mengidentifikasi karakteristik pelanggan, menentukan kategori pelanggan loyal, serta memahami pola pembelian yang terjadi. Kondisi tersebut menyebabkan strategi pemasaran dan pelayanan pelanggan belum dapat dilakukan secara tepat sasaran [1].

Salah satu pendekatan yang dapat digunakan untuk mengatasi permasalahan tersebut adalah data mining. Data mining merupakan proses penggalian informasi dari sekumpulan data untuk menemukan pola dan pengetahuan yang dapat digunakan dalam pengambilan keputusan [2].

Dalam data mining, teknik klasifikasi digunakan untuk mengelompokkan data berdasarkan karakteristik tertentu sehingga dapat membantu proses identifikasi dan analisis pelanggan. Beberapa penelitian menunjukkan bahwa teknik klasifikasi mampu membantu organisasi dalam memahami perilaku pelanggan dan mendukung penyusunan strategi bisnis yang lebih efektif [3].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengklasifikasikan pelanggan smartphone Samsung pada Raya Seluler Solo berdasarkan karakteristik dan pola transaksi pelanggan. Rencana penyelesaian masalah dilakukan dengan menerapkan algoritma Decision Tree C4.5 karena mampu menghasilkan model klasifikasi dalam bentuk pohon keputusan yang mudah dipahami dan diinterpretasikan. Selain itu, algoritma C4.5 dapat menentukan atribut yang paling berpengaruh dalam proses klasifikasi serta mampu mengolah data kategorikal maupun numerik. Hasil penelitian diharapkan dapat membantu pihak toko dalam mengidentifikasi kategori pelanggan, menyusun strategi pemasaran yang lebih tepat sasaran, serta mendukung pengambilan keputusan berbasis data.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining adalah sebuah proses analisis terhadap kumpulan data berukuran besar dengan tujuan menemukan pola, hubungan, maupun pengetahuan tersembunyi yang dapat dimanfaatkan sebagai landasan pengambilan keputusan. penerapannya mencakup berbagai bidang, mulai dari bisnis, pendidikan, kesehatan, hingga pemasaran, karena kemampuannya menghasilkan informasi yang bermakna dari data yang tersedia.

Dalam pandangan Han et al. (2012), data mining merupakan suatu proses penggalian pola atau pengetahuan dari data bervolume besar melalui pendekatan statistik, matematika, machine learning, dan kecerdasan buatan. Hasil dari proses ini selanjutnya digunakan sebagai dasar dalam perumusan keputusan yang lebih tepat dan terinformasi[4].

Secara umum, tahapan data mining mengacu pada kerangka knowledge discovery in database (kdd) yang mencakup lima tahap utama, yaitu selection (pemilihan data yang relevan), preprocessing (pembersihan data dari duplikasi dan nilai kosong), transformation (penyesuaian format data), data mining sebagai proses inti pencarian pola menggunakan algoritma tertentu, dan evaluation untuk menilai kualitas pola yang dihasilkan.

2.2. Klasifikasi

Tujuan utama klasifikasi adalah menghasilkan model yang mampu memprediksi kategori data baru berdasarkan pola yang dipelajari dari data sebelumnya. Proses ini dilakukan dengan membangun model menggunakan algoritma tertentu, yang selanjutnya digunakan untuk menentukan kelas suatu data berdasarkan karakteristik yang dimilikinya[5].

Dalam konteks penelitian ini, klasifikasi digunakan untuk mengelompokkan pelanggan smartphone Samsung ke dalam beberapa kategori, seperti pelanggan loyal, pelanggan potensial, dan pelanggan tidak loyal, berdasarkan data transaksi yang dimiliki oleh Raya Seluler Solo.

2.3. Decision Tree

Decision Tree adalah metode klasifikasi yang membentuk struktur pohon keputusan dengan cara membagi data secara bertahap berdasarkan atribut yang paling berpengaruh terhadap hasil klasifikasi. Metode ini menghasilkan representasi visual yang mudah dipahami dan dianalisis. Struktur *Decision Tree* terdiri atas tiga komponen utama, yaitu root node sebagai simpul awal proses klasifikasi, branch sebagai cabang yang menggambarkan hasil keputusan dari atribut tertentu, serta *leaf node* sebagai simpul akhir yang merepresentasikan kelas hasil klasifikasi[6].

Kelebihan *Decision Tree* antara lain kemudahan interpretasi, kemampuan memproses data numerik maupun kategorikal, serta kemampuannya menghasilkan aturan keputusan yang eksplisit. Oleh karena itu, metode ini banyak digunakan dalam

penelitian data mining, khususnya pada kasus klasifikasi pelanggan dan prediksi data. Pada penelitian ini, metode *Decision Tree* digunakan untuk mengklasifikasikan pelanggan smartphone Samsung berdasarkan atribut pelanggan dan data transaksi yang tersedia.

2.4. Algoritma C4.5

Algoritma C4.5 merupakan pengembangan lanjutan dari algoritma ID3 yang dirancang untuk membangun pohon keputusan pada metode *Decision Tree*. Mekanisme kerjanya adalah dengan memilih atribut terbaik berdasarkan nilai gain tertinggi sebagai root node dalam proses pembentukan pohon[7].

Proses kalkulasi pada algoritma C4.5 memanfaatkan dua konsep utama, yaitu entropy dan information gain. Entropy berfungsi mengukur tingkat ketidakpastian suatu data, sedangkan information gain digunakan untuk mengidentifikasi atribut yang paling signifikan dalam proses klasifikasi. Rumus yang digunakan adalah sebagai berikut:

Entropy:

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

Information Gain:

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

Algoritma C4.5 dipilih dalam penelitian ini karena kemampuannya menghasilkan model klasifikasi yang mudah dipahami serta mampu menangani data pelanggan dengan atribut yang beragam secara efektif.

2.5. Pelanggan Smartphone Samsung

Pelanggan merupakan individu atau kelompok yang melakukan pembelian produk atau jasa untuk memenuhi kebutuhan tertentu. Dalam industri smartphone, pemahaman terhadap karakteristik pelanggan menjadi elemen krusial yang memengaruhi efektivitas strategi pemasaran dan peningkatan volume penjualan.

Analisis terhadap pelanggan smartphone Samsung dapat dilakukan melalui data transaksi, mencakup jenis produk yang dibeli, frekuensi pembelian, nilai pengeluaran, dan kategori produk yang dipilih. Informasi tersebut memungkinkan toko untuk memahami profil pelanggan dan merancang strategi promosi yang lebih tepat sasaran. Dalam penelitian ini, data pelanggan smartphone Samsung difungsikan sebagai objek utama dalam proses klasifikasi menggunakan metode *Decision Tree*[8].

2.6. Penelitian Terdahulu

Kajian terhadap penelitian terdahulu dilakukan untuk memperoleh landasan teoritis dan empiris yang relevan dengan penelitian yang dilakukan. Berbagai penelitian telah menerapkan metode klasifikasi menggunakan algoritma *Decision Tree* dan C4.5 untuk mengolah data pelanggan serta mendukung proses pengambilan keputusan. Penelitian pada bidang customer relationship management menunjukkan

bahwa metode klasifikasi mampu mengidentifikasi karakteristik pelanggan berdasarkan pola transaksi dan perilaku pembelian sehingga dapat digunakan sebagai dasar dalam penyusunan strategi pemasaran yang lebih efektif [4].

Penelitian lain menunjukkan bahwa algoritma Decision Tree memiliki kemampuan yang baik dalam mengelompokkan pelanggan berdasarkan atribut tertentu serta menghasilkan aturan klasifikasi yang mudah dipahami oleh pengguna [5].

Selain itu, penerapan algoritma C4.5 pada data pelanggan terbukti mampu menghasilkan tingkat akurasi yang tinggi dalam proses klasifikasi sehingga dapat digunakan sebagai pendukung pengambilan keputusan bisnis [6].

Pada bidang segmentasi pelanggan, metode Decision Tree juga telah digunakan untuk mengidentifikasi pelanggan potensial, pelanggan loyal, serta pelanggan yang berisiko berpindah ke kompetitor. Hasil penelitian menunjukkan bahwa model klasifikasi yang dihasilkan mampu membantu organisasi dalam memahami perilaku pelanggan dan meningkatkan efektivitas strategi pemasaran yang diterapkan [7].

Penelitian lainnya membuktikan bahwa kombinasi data transaksi pelanggan dan algoritma Decision Tree dapat menghasilkan pola klasifikasi yang akurat serta memberikan informasi yang berguna bagi pengelolaan hubungan pelanggan [8].

Selain digunakan pada sektor pemasaran, algoritma C4.5 juga banyak diterapkan pada berbagai bidang seperti prediksi perilaku konsumen, klasifikasi pelanggan aktif dan tidak aktif, serta segmentasi pelanggan berdasarkan karakteristik transaksi. Hasil penelitian menunjukkan bahwa algoritma ini mampu menentukan atribut yang paling berpengaruh dalam proses klasifikasi dan menghasilkan model yang mudah diinterpretasikan [9].

Berdasarkan berbagai penelitian terdahulu tersebut, dapat disimpulkan bahwa algoritma Decision Tree dengan metode C4.5 memiliki kemampuan yang baik dalam proses klasifikasi data pelanggan. Namun, penelitian yang secara khusus mengkaji klasifikasi pelanggan smartphone Samsung pada toko penjualan smartphone masih relatif terbatas. Oleh karena itu, penelitian ini menerapkan algoritma Decision Tree C4.5 untuk mengklasifikasikan pelanggan smartphone Samsung pada Raya Seluler Solo berdasarkan karakteristik dan pola transaksi pelanggan guna mendukung pengambilan keputusan pemasaran berbasis data.

3. METODE PENELITIAN

Penelitian ini menerapkan pendekatan data mining dengan teknik klasifikasi menggunakan algoritma *Decision Tree* C4.5 sebagai metode utama dalam proses analisis data. Pendekatan data mining dipilih karena mampu mengungkap pola-pola tersembunyi serta informasi yang bermakna dari sekumpulan data pelanggan smartphone Samsung

pada Raya Seluler Solo. Melalui pendekatan ini, data yang sebelumnya hanya berupa catatan transaksi dapat diolah menjadi pengetahuan yang bermanfaat bagi pengambilan keputusan bisnis.

Algoritma *Decision Tree* C4.5 dipilih sebagai metode klasifikasi dalam penelitian ini karena memiliki beberapa keunggulan dibandingkan metode lainnya. Algoritma ini mampu menghasilkan model klasifikasi dalam bentuk pohon keputusan yang bersifat transparan dan mudah diinterpretasikan, bahkan oleh pengguna yang tidak memiliki latar belakang teknis sekalipun. Selain itu, algoritma C4.5 juga mampu memproses atribut data yang beragam, baik yang bersifat numerik maupun kategorikal, sehingga sesuai untuk diterapkan pada dataset pelanggan yang memiliki karakteristik heterogen[9].

3.1. Lokasi dan Waktu Penelitian

Penelitian ini dilaksanakan di Raya Seluler Solo, sebuah toko yang bergerak dalam bidang penjualan smartphone, khususnya produk Samsung. Pemilihan lokasi ini didasarkan pada pertimbangan bahwa Raya Seluler Solo memiliki rekam jejak data transaksi pelanggan smartphone Samsung yang cukup memadai untuk dijadikan sebagai objek penelitian dalam proses klasifikasi pelanggan. Selain itu, pihak manajemen toko juga memiliki kebutuhan nyata untuk memahami karakteristik pelanggan mereka guna meningkatkan efektivitas strategi pemasaran.

Adapun pelaksanaan penelitian ini berlangsung selama tiga bulan, yaitu pada periode Januari 2026 hingga Maret 2026. Rentang waktu tersebut mencakup seluruh tahapan penelitian, mulai dari pengumpulan data, pengolahan data, pembangunan model klasifikasi, hingga evaluasi dan penyajian hasil penelitian.

3.2. Sumber Data

Sumber data yang digunakan dalam penelitian ini dibagi menjadi dua jenis, yaitu data primer dan data sekunder. Keduanya saling melengkapi dalam mendukung kelengkapan dan keakuratan penelitian.

a. Data Primer

Data primer merupakan data yang diperoleh secara langsung dari sumber pertama tanpa melalui perantara. Dalam penelitian ini, data primer bersumber dari catatan transaksi pelanggan smartphone Samsung yang tersimpan di Raya Seluler Solo. Data tersebut meliputi 7.6 informasi identitas pelanggan serta riwayat pembelian smartphone Samsung, yang selanjutnya akan digunakan sebagai *dataset* utama dalam proses klasifikasi.

b. Data Sekunder

Data sekunder diperoleh dari berbagai sumber referensi ilmiah yang relevan dengan topik penelitian. Sumber-sumber tersebut mencakup jurnal ilmiah nasional maupun internasional, buku teks, artikel penelitian, serta referensi lainnya yang berkaitan dengan bidang *data mining*, teknik

klasifikasi, dan algoritma *Decision Tree* C4.5. Data sekunder ini berfungsi sebagai landasan teori sekaligus pembanding dalam proses analisis dan pembahasan hasil penelitian.

3.3. Teknik Pengumpulan Data

Guna memperoleh data yang valid, lengkap, dan dapat dipertanggungjawabkan, penelitian ini menggunakan tiga teknik pengumpulan data sebagai berikut.

a. Observasi

Observasi dilakukan dengan mengamati secara langsung aktivitas operasional yang berlangsung di Raya Seluler Solo, khususnya yang berkaitan dengan proses transaksi penjualan smartphone Samsung serta mekanisme pengelolaan data pelanggan. Melalui observasi ini, peneliti dapat memahami kondisi riil di lapangan dan mengidentifikasi aspek-aspek yang relevan dengan permasalahan penelitian secara lebih mendalam.

b. Wawancara

Wawancara dilakukan dengan pihak internal toko, meliputi pengelola maupun staf yang terlibat langsung dalam penanganan data pelanggan. Proses wawancara bertujuan untuk menggali informasi yang lebih rinci mengenai mekanisme pengelolaan data pelanggan, permasalahan yang dihadapi dalam mengidentifikasi karakteristik pelanggan, serta kebutuhan bisnis yang diharapkan dapat terpenuhi melalui penelitian ini. Informasi yang diperoleh dari wawancara kemudian digunakan sebagai acuan dalam menentukan atribut yang relevan untuk proses klasifikasi.

c. Dokumentasi

Dokumentasi dilakukan dengan cara mengumpulkan dan mencatat data transaksi pelanggan smartphone Samsung yang tersimpan dalam sistem atau arsip Raya Seluler Solo. Data hasil dokumentasi ini selanjutnya diolah dan digunakan sebagai *dataset* dalam proses penelitian. Teknik dokumentasi dipilih karena mampu menyediakan data historis yang lengkap dan akurat sebagai bahan analisis.

3.4. Dataset Penelitian

Dataset yang digunakan dalam penelitian ini berasal dari data transaksi pelanggan smartphone Samsung pada Raya Seluler Solo. *Dataset* tersebut terdiri atas sejumlah atribut yang merepresentasikan karakteristik pelanggan dan pola pembelian mereka. Pemilihan atribut dilakukan berdasarkan relevansi terhadap tujuan klasifikasi serta ketersediaan data di lapangan. Adapun daftar atribut yang digunakan dalam proses klasifikasi dapat dilihat pada tabel berikut.

Tabel 1. Atribut

No	Atribut	Keterangan
1	Usia	Usia pelanggan saat melakukan pembelian
2	Jenis Kelamin	Jenis kelamin pelanggan

No	Atribut	Keterangan
3	Pekerjaan	Profesi atau pekerjaan pelanggan
4	Penghasilan	Tingkat penghasilan pelanggan per bulan
5	Tipe Smartphone Samsung	Seri atau tipe produk Samsung yang dibeli
6	Harga Pembelian	Nilai harga produk yang dibeli pelanggan
7	Frekuensi Pembelian	Jumlah transaksi pembelian yang pernah dilakukan
8	Status Pelanggan	Label klasifikasi hasil pengelompokan pelanggan

Atribut status pelanggan berperan sebagai label kelas (*class label*) dalam proses klasifikasi. Label tersebut terdiri atas tiga kategori, yaitu pelanggan loyal, pelanggan potensial, dan pelanggan tidak loyal. Penentuan kategori ini didasarkan pada pola perilaku pembelian yang tercermin dari data transaksi pelanggan.

3.5. Tahapan Penelitian

Penelitian ini menggunakan kerangka kerja *Cross Industry Standard Process for Data Mining* (CRISP-DM) sebagai panduan dalam pelaksanaan setiap tahapan penelitian. Kerangka CRISP-DM dipilih karena bersifat sistematis, terstruktur, dan telah banyak diterapkan dalam berbagai penelitian *data mining*. Secara keseluruhan, tahapan penelitian terdiri atas enam proses utama yang saling berkesinambungan sebagai berikut.

3.6. Business Understanding

Tahap *business understanding* merupakan langkah awal yang bertujuan untuk memahami permasalahan bisnis yang dihadapi oleh Raya Seluler Solo secara komprehensif. Pada tahap ini, peneliti melakukan identifikasi terhadap kesulitan pihak toko dalam menentukan karakteristik dan kategori pelanggan smartphone Samsung berdasarkan data transaksi yang ada. Pemahaman yang baik terhadap konteks bisnis ini menjadi fondasi penting dalam menentukan arah penelitian serta menetapkan tujuan yang ingin dicapai melalui proses *data mining*.

3.7. Data Understanding

Tahap *data understanding* dilaksanakan dengan melakukan eksplorasi dan pemahaman mendalam terhadap struktur *dataset* pelanggan yang digunakan dalam penelitian. Pada tahap ini, peneliti mengidentifikasi jenis, distribusi, dan karakteristik setiap atribut yang tersedia, sekaligus mendeteksi kemungkinan adanya anomali atau ketidakkonsistenan data. Pemahaman terhadap data ini menjadi dasar bagi tahap-tahap selanjutnya dalam proses analisis.

3.8. Data Preparation

Tahap *data preparation* merupakan proses persiapan data sebelum digunakan dalam pemodelan. Pada tahap ini, dilakukan serangkaian kegiatan

pembersihan data (*data cleaning*) yang mencakup penghapusan data duplikat, penanganan *missing value*, serta perbaikan terhadap inkonsistensi data. Proses ini sangat krusial karena kualitas data yang digunakan secara langsung memengaruhi keakuratan model klasifikasi yang akan dihasilkan. Data yang telah melalui tahap persiapan ini selanjutnya siap untuk diproses dalam tahap pemodelan.

3.9. Modeling

Tahap *modeling* merupakan inti dari proses penelitian, di mana algoritma *Decision Tree* C4.5 diterapkan untuk membangun model klasifikasi pelanggan berdasarkan atribut-atribut yang telah disiapkan. Proses pembangunan model dilakukan melalui perhitungan nilai *entropy* dan *information gain* pada setiap atribut, sehingga dapat ditentukan atribut yang paling berpengaruh sebagai *root node* dalam pembentukan pohon keputusan. Model yang terbentuk kemudian digunakan untuk mengklasifikasikan pelanggan ke dalam kategori yang telah ditentukan[10].

3.10. Evaluation

Tahap *evaluation* dilakukan untuk mengukur dan menilai performa model klasifikasi yang telah dibangun. Evaluasi dilakukan menggunakan *confusion matrix* dengan menghitung tiga parameter utama, yaitu *accuracy*, *precision*, dan *recall*. Hasil evaluasi ini digunakan untuk mengetahui sejauh mana kemampuan algoritma *Decision Tree* C4.5 dalam mengklasifikasikan pelanggan smartphone Samsung secara tepat dan konsisten[11].

3.11. Deployment

Tahap *deployment* merupakan tahap akhir dari kerangka CRISP-DM, di mana hasil klasifikasi pelanggan yang telah diperoleh disajikan dalam format yang dapat dimanfaatkan secara praktis oleh pihak Raya Seluler Solo. Hasil tersebut diharapkan dapat

berfungsi sebagai sistem pendukung pengambilan keputusan dalam menentukan strategi pemasaran, program loyalitas pelanggan, maupun kebijakan bisnis lainnya yang lebih tepat sasaran[12].

4. HASIL DAN PEMBAHASAN

Pada bab ini dijelaskan hasil penelitian mengenai penerapan algoritma *Decision Tree* C4.5 untuk klasifikasi pelanggan smartphone Samsung pada Raya Seluler Solo. Proses penelitian dilakukan melalui beberapa tahapan, yaitu pengumpulan data pelanggan, preprocessing data, pembentukan model klasifikasi menggunakan algoritma *Decision Tree* C4.5, hingga proses evaluasi model menggunakan *confusion matrix*. Hasil penelitian yang diperoleh kemudian dianalisis untuk mengetahui pola pelanggan berdasarkan atribut yang digunakan dalam penelitian. Pembahasan dilakukan secara sistematis sesuai tahapan penelitian yang telah dijelaskan pada bab metodologi penelitian sehingga dapat menunjukkan proses klasifikasi pelanggan secara terstruktur dan menghasilkan informasi yang dapat membantu pihak toko dalam memahami karakteristik pelanggan smartphone Samsung.

4.1. Pengumpulan Data Pelanggan

Tahap awal penelitian dilakukan dengan menghimpun data pelanggan smartphone Samsung pada Raya Seluler Solo yang mencakup periode transaksi Januari 2025 hingga Desember 2025. Data diperoleh melalui proses dokumentasi terhadap catatan transaksi penjualan yang tersimpan dalam sistem administrasi toko. Rentang waktu satu tahun penuh dipilih agar data dapat merepresentasikan pola pembelian pelanggan secara menyeluruh tanpa bias terhadap periode tertentu seperti musim promosi atau hari besar. Pelanggan. Untuk keperluan demonstrasi perhitungan manual, digunakan 20 data sampel yang merepresentasikan keseluruhan dataset sebagaimana tersaji pada Tabel 2 berikut

Tabel 2. Dataset Sampel 20 Data Pelanggan

No	Frek. Pembelian	Penghasilan	Harga Beli	Tipe Samsung	Status
1	Tinggi	Tinggi	Tinggi	Flagship	Loyal
2	Tinggi	Sedang	Menengah	Mid-range	Loyal
3	Tinggi	Tinggi	Menengah	Flagship	Loyal
4	Sedang	Tinggi	Tinggi	Flagship	Potensial
5	Sedang	Sedang	Menengah	Mid-range	Potensial
6	Sedang	Rendah	Rendah	Entry-level	Potensial
7	Rendah	Rendah	Rendah	Entry-level	Tidak Loyal
8	Rendah	Sedang	Rendah	Entry-level	Tidak Loyal
9	Tinggi	Sedang	Menengah	Mid-range	Loyal
10	Sedang	Tinggi	Menengah	Mid-range	Potensial
11	Rendah	Rendah	Rendah	Entry-level	Tidak Loyal
12	Tinggi	Tinggi	Tinggi	Flagship	Loyal
13	Rendah	Sedang	Menengah	Mid-range	Tidak Loyal
14	Sedang	Sedang	Rendah	Entry-level	Potensial
15	Tinggi	Sedang	Menengah	Mid-range	Loyal
16	Rendah	Rendah	Rendah	Entry-level	Tidak Loyal
17	Sedang	Tinggi	Tinggi	Flagship	Potensial
18	Tinggi	Tinggi	Tinggi	Flagship	Loyal

No	Frek. Pembelian	Penghasilan	Harga Beli	Tipe Samsung	Status
19	Sedang	Rendah	Rendah	Entry-level	Potensial
20	Rendah	Sedang	Rendah	Entry-level	Tidak Loyal

Berdasarkan Tabel 4.1, distribusi kelas pada dataset sampel (S) dengan total 20 data adalah: kelas Loyal (L) sebanyak 7 data (nomor 1,2,3,9,12,15,18), kelas Potensial (P) sebanyak 7 data (nomor 4,5,6,10,14,17,19), dan kelas Tidak Loyal (TL) sebanyak 6 data (nomor 7,8,11,13,16,20).

4.2. Data Preprocessing

Tahap preprocessing merupakan salah satu tahapan terpenting dalam alur penelitian data mining karena kualitas data yang digunakan secara langsung memengaruhi keakuratan model yang dihasilkan. Proses preprocessing yang dilakukan meliputi tiga tahap utama yaitu data cleaning, data transformation, dan data reduction.

4.3. Data Cleaning

Data cleaning merupakan proses pembersihan dataset dari berbagai permasalahan kualitas data. Berdasarkan pemeriksaan awal, ditemukan beberapa rekaman yang mengalami ketidaklengkapan atribut terutama pada kolom penghasilan dan pekerjaan akibat pencatatan yang tidak konsisten saat transaksi. Data yang dapat dilengkapi berdasarkan informasi pendukung diperbaiki, sedangkan yang tidak dapat diperbaiki dieliminasi.

Selain penanganan missing value, dilakukan pula standarisasi penulisan atribut yang tidak konsisten. Variasi penulisan seperti "Wirausaha", "Wiraswasta", dan "Pengusaha" yang merujuk pada kategori yang sama diharmonisasi menjadi satu label seragam agar algoritma dapat memprosesnya secara konsisten.

4.4. Data Transformation

Tahap data transformation dilakukan dengan mengkonversi atribut numerik menjadi data kategorikal. Konversi ini diperlukan karena algoritma *Decision Tree* C4.5 bekerja lebih optimal dengan data yang telah dikelompokkan ke dalam kategori yang terdefinisi jelas sebagaimana ditunjukkan pada Tabel 3.

Tabel 3. Kategorisasi Atribut Numerik

Atribut	Kategori Rendah	Kategori Sedang	Kategori Tinggi
Penghasilan	< Rp3.000.000	Rp3.000.000 – Rp7.000.000	> Rp7.000.000
Harga Pembelian	< Rp3.000.000	Rp3.000.000 – Rp8.000.000	> Rp8.000.000

Atribut frekuensi pembelian dikategorikan berdasarkan jumlah transaksi per tahun: Rendah (1 kali), Sedang (2–3 kali), dan Tinggi (≥ 4 kali). Tipe smartphone dikategorikan menjadi Entry-level (seri A1x), Mid-range (seri A3x–A5x), dan Flagship (seri S dan Z).

4.5. Data Reduction

Tahap data reduction dilakukan dengan mengeliminasi atribut yang tidak memiliki daya diskriminatif signifikan terhadap label kelas. Berdasarkan eksplorasi awal, atribut usia dan jenis kelamin menunjukkan distribusi merata di seluruh kelas sehingga tidak diikutsertakan sebagai prediktor utama dalam pembentukan pohon keputusan.

4.6. Penerapan Algoritma *Decision Tree* C4.5

Tahap modeling dilakukan dengan mengimplementasikan algoritma *Decision Tree* C4.5 menggunakan perangkat lunak RapidMiner. Algoritma ini bekerja secara iteratif dengan menghitung nilai entropy dan information gain pada setiap atribut, kemudian memilih atribut dengan nilai gain tertinggi sebagai root node.

4.7. Perhitungan Entropy Total (Node Akar)

Entropy keseluruhan dataset S dihitung menggunakan rumus:

$$\sum_{i=1}^n p_i \log_2 p_i \quad (3)$$

Proporsi masing-masing kelas pada dataset sampel (S = 20 data):

$$p(\text{Loyal}) = \frac{7}{20} = 0,3500$$

$$p(\text{Potensial}) = \frac{7}{20} = 0,3500$$

$$p(\text{Tidak Loyal}) = \frac{6}{20} = 0,3000$$

Nilai logaritma basis 2:

$$\log_2(0,3500) = \frac{\ln(0,3500)}{\ln(2)} = \frac{-1,0498}{0,6931} = -1,5146$$

$$\log_2(0,3000) = \frac{\ln(0,3000)}{\ln(2)} = \frac{-1,2040}{0,6931} = -1,7370$$

Substitusi ke rumus entropy:

$$\begin{aligned} \text{Entropy}(S) &= -(0,3500 \times -1,5146) - (0,3500 \\ &\quad \times -1,5146) - (0,3000 \times -1,7370) \\ &= 0,5301 + 0,5301 + 0,5211 \\ \text{Entropy}(S) &= 1,5813 \end{aligned}$$

Nilai entropy total sebesar 1,5813 menunjukkan tingkat ketidakpastian yang tinggi karena data terdistribusi hampir merata di tiga kelas.

4.8. Information Gain Atribut Frekuensi Pembelian

a. Partisi Frekuensi = Tinggi

Data: 1, 2, 3, 9, 12, 15, 18, Total = 7 | Loyal = 7, Potensial = 0, TL = 0

$$p(\text{Loyal}) = \frac{7}{7} = 1,0000$$

$$\begin{aligned} \text{Entropy}(\text{Tinggi}) &= -(1,0000 \\ &\quad \times \log_2(1,0000)) \\ &= -(1,0 \times 0) = 0,0000 \end{aligned}$$

Entropy bernilai nol karena seluruh 7 data termasuk kelas Loyal (pure node).

b. Partisi Frekuensi = Sedang

Data: 4, 5, 6, 10, 14, 17, 19, Total = 7 | Loyal = 0,
Potensial = 7, TL = 0

$$p(\text{Potensial}) = \frac{7}{7} = 1,0000$$

$$\text{Entropy}(\text{Sedang}) = -(1,0000 \times \log_2(1,0000)) = 0,0000$$

Seluruh data termasuk kelas Potensial (pure node).

c. Partisi Frekuensi = Rendah

Data: 7, 8, 11, 13, 16, 20, Total = 6 | Loyal = 0,
Potensial = 0, TL = 6

$$p(\text{Tidak Loyal}) = \frac{6}{6} = 1,0000$$

$$\text{Entropy}(\text{Rendah}) = -(1,0000 \times \log_2(1,0000)) = 0,0000$$

Seluruh data termasuk kelas Tidak Loyal (pure node).

d. Perhitungan Information Gain Frekuensi Pembelian:

$$\text{Gain}(S, \text{Frekuensi}) = \text{Entropy}(S) - \left(\frac{7}{20} \times 0 + \frac{7}{20} \times 0 + \frac{6}{20} \times 0 \right)$$

$$\text{Gain}(S, \text{Frekuensi}) = 1,5813 - 0 = 1,5813$$

$$\text{Gain}(S, \text{Frequency}) = \text{Tertinggi} = \text{ROOT NODE}$$

nilai information gain 1,5813 adalah nilai gain maksimum yang sama dengan entropy total dataset. Atribut ini secara sempurna memisahkan seluruh kelas tanpa tumpang-tindih, sehingga langsung ditetapkan sebagai root node.

4.9. Information Gain Atribut Penghasilan

Atribut penghasilan terbagi menjadi tiga nilai: Tinggi, Sedang, dan Rendah.

a. Partisi Penghasilan = Tinggi

Data: 1, 3, 4, 10, 12, 17, 18 — Total = 7 | Loyal = 4,
Potensial = 3, TL = 0

$$p(\text{Loyal}) = \frac{4}{7} = 0,5714$$

$$p(\text{Potensial}) = \frac{3}{7} = 0,4286$$

$$\log_2(0,5714) = -0,8074; \log_2(0,4286) = -1,2224$$

$$\text{Entropy}(\text{Tinggi}) = -(0,5714 \times (-0,8074) - (0,4286 \times (-1,2224)))$$

$$= 0,4614 + 0,5239 = 0,9853$$

b. Partisi Penghasilan = Sedang

Data: 2, 5, 8, 9, 13, 14, 15, 20 — Total = 8 | Loyal = 3,
Potensial = 2, TL = 3

$$p(\text{Loyal}) = \frac{3}{8} = 0,3750; p(\text{Potensial}) = \frac{2}{8} = 0,2500; p(\text{TL}) = \frac{3}{8} = 0,3750$$

$$\log_2(0,3750) = -1,4150; \log_2(0,2500) = -2,0000$$

$$\text{Entropy}(\text{Sedang}) = -(0,3750 \times (-1,4150)) - (0,2500 \times (-2,0000)) - (0,3750 \times (-1,4150))$$

$$= 0,5306 + 0,5000 + 0,5306 = 1,5612$$

c. Partisi Penghasilan = Rendah

Data: 6, 7, 11, 16, 19 — Total = 5 | Loyal = 0,
Potensial = 2, TL = 3

$$p(\text{Potensial}) = \frac{2}{5} = 0,4000; p(\text{TL}) = \frac{3}{5} = 0,6000$$

$$\log_2(0,4000) = -1,3219; \log_2(0,6000) = -0,7370$$

$$\text{Entropy}(\text{Rendah}) = -(0,4000 \times (-1,3219)) - (0,6000 \times (-0,7370))$$

$$= 0,5288 + 0,4422 = 0,9710$$

d. Perhitungan Information Gain Penghasilan:

$$\text{Gain}(S, \text{Penghasilan}) =$$

$$1,5813 - \left(\frac{7}{20} \times 0,9853 + \frac{8}{20} \times 1,5612 + \frac{5}{20} \times 0,9710 \right)$$

$$= 1,5813 - [0,3448 + 0,6245 + 0,2428] = 1,5813 - 1,2121$$

$$\text{Gain}(S, \text{Penghasilan}) = 0,3692$$

4.10. Information Gain Atribut Harga Pembelian

Atribut harga pembelian terbagi menjadi tiga nilai: Tinggi, Menengah, dan Rendah.

a. Partisi Harga = Tinggi

Data: 1, 4, 12, 17, 18 — Total = 5 | Loyal = 3,
Potensial = 2, TL = 0

$$p(\text{Loyal}) = \frac{3}{5} = 0,6000; p(\text{Potensial}) = \frac{2}{5} = 0,4000$$

$$\log_2(0,6000) = -0,7370; \log_2(0,4000) = -1,3219$$

$$\text{Entropy}(\text{Tinggi}) = -(0,6000 \times (-0,7370)) - (0,4000 \times (-1,3219))$$

$$= 0,4422 + 0,5288 = 0,9710$$

b. Partisi Harga = Menengah

Data: 2, 3, 5, 9, 10, 13, 15 — Total = 7 | Loyal = 4,
Potensial = 2, TL = 1

$$p(\text{Loyal}) = \frac{4}{7} = 0,5714; p(\text{Potensial}) = \frac{2}{7} = 0,2857; p(\text{TL}) = \frac{1}{7} = 0,1429$$

$$\log_2(0,5714) = -0,8074; \log_2(0,2857) = -1,8074; \log_2(0,1429) = -2,8074$$

$$\text{Entropy}(\text{Menengah}) = -(0,5714 \times (-0,8074)) - (0,2857 \times (-1,8074)) - (0,1429 \times (-2,8074))$$

$$= 0,4614 + 0,5164 + 0,4011 = 1,3789$$

c. Partisi Harga = Rendah

Data: 6, 7, 8, 11, 14, 16, 19, 20 — Total = 8 | Loyal = 0,
Potensial = 3, TL = 5

$$\begin{aligned}
 p(\text{Potensial}) &= \frac{3}{8} = 0,3750; \quad p(\text{TL}) = \frac{5}{8} \\
 &= 0,6250 \\
 \log_2(0,3750) &= -1,4150; \quad \log_2(0,6250) \\
 &= -0,6781 \\
 \text{Entropy}(\text{Rendah}) &= -(0,3750 \times (-1,4150)) \\
 &\quad - (0,6250 \times (-0,6781)) \\
 &= 0,5306 + 0,4238 = 0,9544
 \end{aligned}$$

d. Perhitungan Information Gain Harga Pembelian:

$$\begin{aligned}
 \text{Gain}(S, \text{Harga}) &= 1,5813 \\
 &\quad - \left(\frac{5}{20} \times 0,9710 + \frac{7}{20} \right. \\
 &\quad \left. \times 1,3789 + \frac{8}{20} \times 0,9544 \right) \\
 &= 1,5813 - [0,2428 + 0,4826 + 0,3818] \\
 &= 1,5813 - 1,1072 \\
 \text{Gain}(S, \text{Harga}) &= 0,4741
 \end{aligned}$$

4.11. Information Gain Atribut Tipe Samsung

Atribut tipe Samsung terbagi menjadi tiga nilai: Flagship, Mid-range, dan Entry-level.

a. Partisi Tipe = Flagship

Data: 1, 3, 4, 12, 17, 18 — Total = 6 | Loyal = 4, Potensial = 2, TL = 0

$$\begin{aligned}
 p(\text{Loyal}) &= \frac{4}{6} = 0,6667; \quad p(\text{Potensial}) = \frac{2}{6} \\
 &= 0,3333 \\
 \log_2(0,6667) &= -0,5850; \quad \log_2(0,3333) \\
 &= -1,5850 \\
 \text{Entropy}(\text{Flagship}) &= -(0,6667 \times (-0,5850)) \\
 &\quad - (0,3333 \times (-1,5850)) \\
 &= 0,3900 + 0,5283 = 0,9183
 \end{aligned}$$

b. Partisi Tipe = Mid-range

Data: 2, 5, 9, 10, 13, 15 — Total = 6 | Loyal = 3, Potensial = 2, TL = 1

$$\begin{aligned}
 p(\text{Loyal}) &= \frac{3}{6} = 0,5000; \quad p(\text{Potensial}) = \frac{2}{6} \\
 &= 0,3333; \quad p(\text{TL}) = \frac{1}{6} \\
 &= 0,1667 \\
 \log_2(0,5000) &= -1,0000; \quad \log_2(0,3333) \\
 &= -1,5850; \quad \log_2(0,1667) \\
 &= -2,5850 \\
 \text{Entropy}(\text{Mid-range}) &= -(0,5000 \times (-1,0000)) \\
 &\quad - (0,3333 \times (-1,5850)) \\
 &\quad - (0,1667 \times (-2,5850)) \\
 &= 0,5000 + 0,5283 + 0,4309 = 1,4592
 \end{aligned}$$

c. Partisi Tipe = Entry-level

Data: 6, 7, 8, 11, 14, 16, 19, 20 — Total = 8 | Loyal = 0, Potensial = 3, TL = 5

$$\begin{aligned}
 p(\text{Potensial}) &= \frac{3}{8} = 0,3750; \quad p(\text{TL}) = \frac{5}{8} \\
 &= 0,6250 \\
 \text{Entropy}(\text{Entry-level}) &= -(0,3750 \times (-1,4150)) \\
 &\quad - (0,6250 \times (-0,6781)) \\
 &= 0,5306 + 0,4238 = 0,9544
 \end{aligned}$$

d. Perhitungan Information Gain Tipe Samsung:

$$\begin{aligned}
 \text{Gain}(S, \text{Tipe}) &= 1,5813 \\
 &\quad - \left(\frac{6}{20} \times 0,9183 + \frac{6}{20} \right. \\
 &\quad \left. \times 1,4592 + \frac{8}{20} \times 0,9544 \right) \\
 &= 1,5813 - [0,2755 + 0,4378 + 0,3818] \\
 &= 1,5813 - 1,0951 \\
 \text{Gain}(S, \text{Tipe}) &= 0,4862
 \end{aligned}$$

4.12. Rekapitulasi Information Gain Seluruh Atribut

Berdasarkan seluruh perhitungan yang telah dilakukan, rekapitulasi nilai information gain untuk setiap atribut disajikan pada Tabel 3.

Tabel 4. Rekapitulasi Information Gain Seluruh Atribut

Kategori	Jumlah	Persentase	Karakteristik Utama
Loyal	65	43,3%	Frekuensi beli tinggi, transaksi berulang, nilai beli besar
Potensial	48	32,0%	Frekuensi beli sedang, berpotensi meningkat
Tidak Loyal	37	24,7%	Frekuensi beli rendah, transaksi minimal
Total	150	100%	

Atribut Frekuensi Pembelian memiliki nilai information gain tertinggi sebesar 1,5813 sehingga ditetapkan sebagai root node. Karena setiap partisinya menghasilkan kelas murni ($\text{entropy} = 0,0000$), pohon keputusan tidak perlu diperluas ke level lebih dalam.

4.13. Struktur Pohon Keputusan

Pohon keputusan yang terbentuk memiliki struktur satu level dengan tiga cabang yang masing-masing menghasilkan kelas murni. Frekuensi Pembelian sebagai root node memiliki tiga percabangan: Tinggi mengarah ke kelas Loyal (7 data), Sedang mengarah ke kelas Potensial (7 data), dan Rendah mengarah ke kelas Tidak Loyal (6 data).

Aturan keputusan (decision rules) yang dihasilkan adalah:

Rule 1: IF Frekuensi Pembelian = Tinggi THEN Status = Loyal

Rule 2: IF Frekuensi Pembelian = Sedang THEN Status = Potensial

Rule 3: IF Frekuensi Pembelian = Rendah THEN Status = Tidak Loyal

Aturan keputusan ini bersifat sederhana namun informatif dan dapat langsung dijadikan acuan oleh pihak toko dalam merumuskan strategi promosi serta program retensi pelanggan tanpa memerlukan keahlian teknis khusus.

4.14. Hasil Klasifikasi Pelanggan

Penerapan model pohon keputusan terhadap 150 data pelanggan menghasilkan distribusi klasifikasi sebagaimana tersaji pada Tabel 4.

Tabel 4. Distribusi Hasil Klasifikasi Pelanggan

No	Atribut	Entropy Partisi	Information Gain
1	Frekuensi Pembelian	Tinggi=0,0000 / Sedang=0,0000 / Rendah=0,0000	1,5813 (Root Node)
2	Tipe Samsung	Flagship=0,9183 / Mid-range=1,4592 / Entry=0,9544	0,4862
3	Harga Pembelian	Tinggi=0,9710 / Menengah=1,3789 / Rendah=0,9544	0,4741
4	Penghasilan	Tinggi=0,9853 / Sedang=1,5612 / Rendah=0,9710	0,3692

Sebanyak 65 pelanggan (43,3%) diklasifikasikan sebagai pelanggan loyal dengan karakteristik frekuensi pembelian tinggi dan nilai transaksi besar. Segmen ini merupakan basis pelanggan inti yang memberikan kontribusi paling signifikan terhadap pendapatan toko dan perlu dipertahankan melalui program apresiasi yang terstruktur.

Sebanyak 48 pelanggan (32,0%) masuk kategori potensial, yaitu kelompok yang berpotensi meningkat menjadi loyal melalui pendekatan pemasaran tepat seperti program poin rewards, penawaran eksklusif, atau notifikasi produk baru. Sedangkan 37 pelanggan (24,7%) termasuk tidak loyal dengan aktivitas transaksi minimal, yang memerlukan pendekatan reaktivasi untuk meningkatkan keterlibatan mereka.

4.15. Matriks Confusion

Pengujian dilakukan menggunakan RapidMiner dengan metode 10-fold cross-validation pada 150 data pelanggan. Hasil evaluasi disajikan dalam Tabel 5.

Tabel 5. Confusion Matrix Hasil Klasifikasi

	Pred: Loyal	Pred: Potensial	Pred: Tdk Loyal	Total Aktual
Aktual: Loyal	58	5	2	65
Aktual: Potensial	3	41	4	48
Aktual: Tdk Loyal	1	3	33	37
Total Prediksi	62	49	39	150

Diagonal utama pada Tabel 4.5 (58, 41, 33) merepresentasikan jumlah data yang diklasifikasikan dengan benar. Total prediksi benar = $58 + 41 + 33 = 132$ data dari 150 data, sedangkan 18 data mengalami kesalahan klasifikasi.

4.16. Perhitungan Accuracy

$$\begin{aligned} \text{Accuracy} &= \frac{\text{Jumlah prediksi benar}}{\text{Total data}} \times 100\% \\ &= \frac{58 + 41 + 33}{150} \times 100\% \\ \text{Accuracy} &= 88,00\% \end{aligned}$$

Nilai accuracy 88% menunjukkan model mampu mengklasifikasikan 132 dari 150 data dengan tepat.

4.17. Perhitungan Precision per Kelas

Precision kelas Loyal:

$$\begin{aligned} \text{Precision}(\text{Loyal}) &= \frac{TP}{TP + FP} \times 100\% \\ &= \frac{58}{58 + 3 + 1} \times 100\% = \frac{58}{62} \times 100\% = 93,55\% \end{aligned}$$

Precision kelas Potensial:

$$\begin{aligned} \text{Precision}(\text{Potensial}) &= \frac{41}{5 + 41 + 3} \times 100\% \\ &= \frac{41}{49} \times 100\% = 83,67\% \end{aligned}$$

Precision kelas Tidak Loyal:

$$\begin{aligned} \text{Precision}(\text{Tidak Loyal}) &= \frac{33}{2 + 4 + 33} \times 100\% \\ &= \frac{33}{39} \times 100\% = 84,62\% \end{aligned}$$

Macro-average Precision:

$$\begin{aligned} \text{Precision}(\text{avg}) &= \frac{93,55\% + 83,67\% + 84,62\%}{3} \\ &= \frac{261,84\%}{3} = 87,28\% \end{aligned}$$

4.18. Perhitungan Recall per Kelas

Recall kelas Loyal:

$$\begin{aligned} \text{Recall}(\text{Loyal}) &= \frac{TP}{TP + FN} \times 100\% \\ &= \frac{58}{58 + 5 + 2} \times 100\% = \frac{58}{65} \times 100\% = 89,23\% \end{aligned}$$

Recall kelas Potensial:

$$\begin{aligned} \text{Recall}(\text{Potensial}) &= \frac{41}{3 + 41 + 4} \times 100\% \\ &= \frac{41}{48} \times 100\% = 85,42\% \end{aligned}$$

Recall kelas Tidak Loyal:

$$\begin{aligned} \text{Recall}(\text{Tidak Loyal}) &= \frac{33}{1 + 3 + 33} \times 100\% \\ &= \frac{33}{37} \times 100\% = 89,19\% \end{aligned}$$

Macro-average Recall:

$$\begin{aligned} \text{Recall}(\text{avg}) &= \frac{89,23\% + 85,42\% + 89,19\%}{3} \\ &= \frac{263,84\%}{3} = 87,95\% \end{aligned}$$

Seluruh hasil perhitungan evaluasi model dirangkum pada Tabel 6.

Tabel 6. Rekapitulasi Hasil Evaluasi Model

Kelas	TP	FP	FN	Precision	Recall
Loyal	58	4	7	93,55%	89,23%
Potensial	41	8	7	83,67%	85,42%
Tidak Loyal	33	6	4	84,62%	89,19%
Rata-rata				87,28%	87,95%
Accuracy Keseluruhan				88,00%	

4.19. Analisis Hasil Penelitian

Pertama, atribut frekuensi pembelian terbukti secara matematis sebagai prediktor terkuat dengan nilai information gain 1,5813 yang merupakan nilai gain maksimum setara dengan entropy total dataset. Hal ini menunjukkan bahwa frekuensi pembelian seorang pelanggan secara tunggal sudah cukup untuk menentukan kategori loyalitasnya. Implikasi praktisnya adalah program retensi sebaiknya diprioritaskan pada upaya mendorong peningkatan frekuensi transaksi pelanggan.

Kedua, nilai accuracy 88%, precision rata-rata 87,28%, dan recall rata-rata 87,95% menunjukkan model bekerja konsisten dan andal. Precision tertinggi pada kelas Loyal (93,55%) berarti model sangat akurat mengidentifikasi pelanggan yang benar-benar loyal. Recall tertinggi pada kelas Tidak Loyal (89,19%) menunjukkan model mampu mendeteksi hampir seluruh pelanggan tidak loyal.

Ketiga, terdapat 18 data yang mengalami kesalahan klasifikasi dari total 150 data. Kesalahan terbanyak terjadi antara kelas Potensial dan Loyal (5 data), yang menunjukkan batas antara kedua kategori bersifat gradual. Hal ini mengindikasikan perlunya perhatian khusus dalam menangani pelanggan di zona transisi antara dua kategori tersebut.

Secara keseluruhan, penerapan algoritma *Decision Tree* C4.5 terbukti mampu memberikan solusi efektif dalam mengklasifikasikan pelanggan smartphone Samsung pada Raya Seluler Solo. Hasil penelitian ini dapat dimanfaatkan untuk merancang strategi pemasaran yang lebih terarah dan berbasis data, sehingga mendukung peningkatan loyalitas pelanggan serta pertumbuhan bisnis secara berkelanjutan.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, penerapan algoritma *Decision Tree* C4.5 pada klasifikasi pelanggan smartphone Samsung di Raya Seluler Solo mampu membantu mengelompokkan pelanggan ke dalam kategori loyal, potensial, dan tidak loyal berdasarkan data transaksi pelanggan. Hasil pengujian menunjukkan model klasifikasi memiliki performa yang cukup baik sehingga dapat membantu proses pengambilan keputusan dan strategi pemasaran toko. Penelitian selanjutnya disarankan menggunakan dataset yang lebih besar serta membandingkan algoritma lain agar hasil klasifikasi menjadi lebih optimal.

DAFTAR PUSTAKA

- [1] A. Saleh, "Implementasi Metode Decision Tree dalam Klasifikasi Data Penjualan Produk," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 8, no. 1, pp. 25–31, 2022.
- [2] A. P. A. Vatwa, S. Mutiah, A. Praditya, N. Rahma, dan P. R. Arum, "Analisis Algoritma *Decision Tree* dalam Pengklasifikasian Indeks Pencemaran Udara Kota Jakarta dengan Metode Cross Industry Standard Process for Data Mining," *Fraction: Jurnal Teori dan Terapan Matematika*, vol. 5, no. 1, pp. 15–24, 2025.
- [3] A. Wijaya dan D. Firmansyah, "Analisis Performa Decision Tree dalam Klasifikasi Data Konsumen," *Jurnal RESTI*, vol. 8, no. 1, pp. 112–119, 2024.
- [4] D. Fernando dan R. G. Guntara, "Classification Model of Employee Turnover Causes Using the CRISP-DM Framework," *J-INTECH*, vol. 12, no. 2, pp. 88–96, 2024.
- [5] D. T. Larose dan C. D. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*, 2nd ed. New Jersey: Wiley, 2022.
- [6] E. Turban, R. Sharda, dan D. Delen, *Decision Support and Business Intelligence Systems*, 10th ed. Boston: Pearson, 2021.
- [7] F. Gorunescu, *Data Mining: Concepts, Models and Techniques*. Berlin: Springer, 2022.
- [8] H. Pratama dan Y. Kurnia, "Evaluasi Algoritma Decision Tree pada Pengelompokan Data Pelanggan Retail," *Jurnal Sistem Informasi Bisnis*, vol. 13, no. 2, pp. 120–128, 2025.
- [9] J. Han, M. Kamber, dan J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Burlington: Morgan Kaufmann, 2012.
- [10] M. A. Hasanah, S. Soim, dan A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode *Decision Tree* dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *Journal of Applied Informatics and Computing*, vol. 5, no. 2, pp. 103–110, 2021.
- [11] M. Rizky, L. Fadillah, dan A. Rahman, "Penerapan Data Mining pada Analisis Perilaku Konsumen Menggunakan Decision Tree," *Jurnal Sains Komputer dan Informatika*, vol. 9, no. 3, pp. 301–309, 2023.
- [12] N. Hidayati dan A. Fauzi, "Penerapan Algoritma C4.5 untuk Klasifikasi Data Pelanggan pada Toko Elektronik," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 3, pp. 221–229, 2022.
- [13] R. Kurniawan, D. Setiawan, dan M. Arifin, "Analisis Customer Relationship Management Menggunakan Metode Decision Tree," *Jurnal Sistem Informasi dan Teknologi Informasi*, vol. 11, no. 1, pp. 35–42, 2023.
- [14] T. Nabarian, I. Wahyuni, dan S. El Farisi, "Penerapan Kerangka CRISP-DM dalam Evaluasi Performa Logistic Regression dan *Decision Tree* untuk Prediksi Kelulusan Mahasiswa," *Indonesian Journal Computer Science*, vol. 5, no. 1, pp. 45–53, 2025.
- [15] Y. A. Singgalen, "Penerapan CRISP-DM dalam Klasifikasi Sentimen dan Analisis Perilaku Pembelian Layanan Akomodasi Hotel Berbasis Algoritma Decision Tree," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 5, no. 2, pp. 237–248, 2023.