



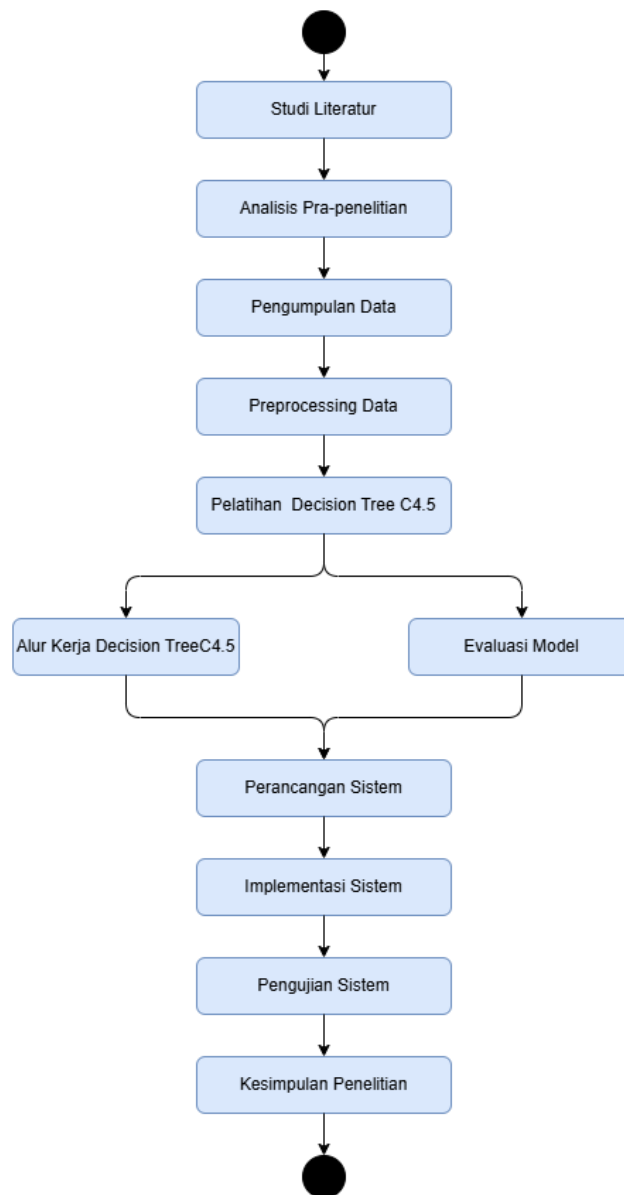
BAB III

METODE PENELITIAN

Bab ini menjelaskan secara rinci pendekatan dan langkah-langkah yang digunakan dalam penelitian untuk mengembangkan sistem rekomendasi smartphone berbasis web menggunakan Decision Tree C4.5. Proses pengembangan dilakukan melalui beberapa tahapan, mulai dari studi literatur, analisis pra-penelitian, pengumpulan data, preprocessing data, pelatihan Decision Tree C4.5, evaluasi Decision Tree C4.5, perancangan sistem, implementasi sistem, hingga pengujian dan evaluasi performa sistem. Setiap tahapan dirancang agar mendukung pencapaian sistem yang akurat dan efisien,

3.1 Kerangka Penelitian

Dalam rangka mendukung pelaksanaan penelitian ini agar lebih terarah dan sistematis, diperlukan suatu kerangka kerja penelitian yang tersusun secara terstruktur. Kerangka kerja ini mencakup tahapan-tahapan penelitian yang bertujuan untuk memberikan panduan dari tahap awal hingga tahap akhir penelitian. Dengan adanya kerangka kerja ini, proses penelitian dapat berjalan secara terorganisasi dan sesuai dengan tujuan yang telah ditetapkan. Tahapan-tahapan tersebut dirancang untuk memastikan bahwa setiap langkah yang dilakukan memberikan kontribusi nyata terhadap pencapaian hasil yang diinginkan. Secara visual, tahapan-tahapan dalam penelitian ini disajikan pada Gambar 3.1.



Gambar 3.1 Kerangka Penelitian

Gambar 3.1, menjelaskan penelitian ini dilakukan melalui tahapan-tahapan sistematis sebagai berikut.

3.1.1 Studi Literatur

Studi literatur dilakukan untuk mengkaji penelitian-penelitian terdahulu yang relevan dengan sistem rekomendasi dan penerapan algoritma Decision Tree C4.5 dalam klasifikasi, serta studi terdahulu terkait tema penelitian ini. Selain itu,

literatur yang dikaji juga mencakup metode evaluasi model klasifikasi menggunakan confusion matrix untuk mengukur akurasi, precision, recall, dan F1-score.

3.1.2 Analisis Pra-Penelitian

Analisis pra-penelitian dilakukan untuk merumuskan masalah dan pendekatan pengembangan sistem. Permasalahan utama yang diidentifikasi adalah kesulitan pelanggan memilih smartphone yang sesuai kebutuhan di tengah banyaknya pilihan produk, serta proses rekomendasi manual oleh tenaga penjual yang memakan waktu dan tidak optimal. Berdasarkan studi literatur dan kondisi lapangan, peneliti memilih algoritma Decision Tree C4.5 karena kesesuaian karakteristiknya dalam menangani data dengan atribut campuran dan menghasilkan pohon keputusan yang interpretatif. Fokus penelitian adalah mengevaluasi akurasi dan menguji efektivitas C4.5 secara spesifik pada toko retail lokal.

3.1.3 Pengumpulan Data

Pengumpulan data dilakukan untuk memperoleh informasi spesifikasi smartphone dan preferensi kebutuhan pengguna sebagai bahan pelatihan model klasifikasi. Data dikumpulkan dari dua sumber, yaitu data primer dari Toko Sahabat Ponsel Tembilahan berupa informasi preferensi dan kebutuhan pelanggan melalui konsultasi dengan tenaga penjual, serta data sekunder berupa spesifikasi smartphone dari dataset publik Kaggle yang diverifikasi dan dilengkapi menggunakan spesifikasi resmi dari GSMArena, dan marketplace Indonesia. Label kategori kebutuhan pengguna (Gaming, Fotografi/Videografi, Pelajar, Pengguna Umum) ditentukan melalui riset artikel dan ulasan terkait kegunaan tiap perangkat, kemudian divalidasi melalui konsultasi dengan tenaga penjual Toko Sahabat Ponsel



Tembilahan untuk memastikan kesesuaian label dengan kondisi kebutuhan pelanggan di lapangan. Data yang telah dikurasi kemudian digabungkan menjadi dataset akhir yang komprehensif untuk pelatihan model.

Tabel 3.1 Pengumpulan Dataset

No	Kelas	Jumlah Data	Persentase
1	Gaming	125	25%
2	Fotografi/Videografi	125	25%
3	Pelajar	125	25%
4	Pengguna Umum	125	25%
Total		500	100%

3.1.4 Preprocessing Data

Preprocessing data dilakukan untuk mengubah data mentah menjadi format yang siap digunakan dalam pelatihan model algoritma C4.5. Data yang digunakan diklasifikasikan ke dalam empat kelas kebutuhan pengguna, yaitu Gaming, Fotografi/Videografi, Pelajar, dan Pengguna Umum. Tahapan preprocessing meliputi:

1. Proses dimulai dengan Exploratory Data Analysis (EDA) untuk mengidentifikasi masalah data seperti missing values, outlier, dan keseimbangan kelas melalui analisis statistik dan visualisasi. Data Cleaning dilakukan dengan mengatasi missing values, menghapus duplikat, memperbaiki inkonsistensi format, dan menangani outlier.
2. Data Integration Data Integration menggabungkan data primer dari Toko Sahabat Ponsel Tembilahan dengan data sekunder dari Kaggle, GSMArena, dan Google. Data Transformation mencakup label encoding untuk atribut



kategorikal. Data Splitting dilakukan secara stratified dengan proporsi 70% training, 15% validation, dan 15% testing untuk menjaga keseimbangan distribusi kelas di setiap subset.

Tabel 3.2 Pembagian Dataset

No	Kelas	Dataset			Total
		Training (70%)	Validasi (15%)	Testing (15%)	
1	Gaming	88	19	18	125
2	Fotografi/Videografi	87	19	19	125
3	Pelajar	88	18	19	125
4	Pengguna Umum	87	19	19	125
Total		350	75	75	500

Metode pembagian data yang digunakan dalam penelitian ini mengaplikasikan Stratified Sampling dengan proporsi 70% : 15% : 15% untuk memastikan proses pelatihan model Decision Tree C4.5 berjalan optimal dan teruji, sambil menjaga keseimbangan distribusi kelas di setiap subset. Pembagian ini menghasilkan tiga himpunan data dengan fungsi yang spesifik: Training Set (70%) digunakan sebagai data utama untuk melatih model algoritma C4.5 hingga terbentuk pohon keputusan; Validation Set (15%) berfungsi untuk mengevaluasi performa model selama proses pelatihan dan berperan penting dalam mencegah overfitting, sementara Testing Set (15%) dialokasikan khusus untuk evaluasi akhir performa model pada data yang belum pernah dilihat, memberikan ukuran akurasi yang objektif terhadap kemampuan generalisasi model.

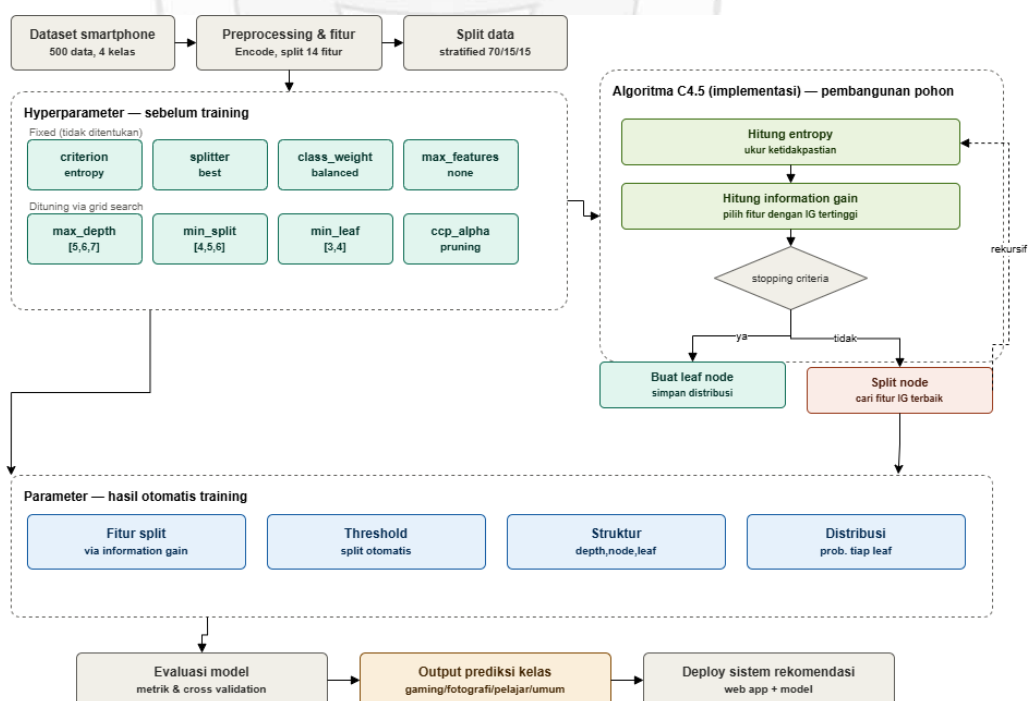


3.1.5 Pelatihan Model C4.5

Tahap pelatihan model merupakan proses inti dalam pengembangan sistem rekomendasi smartphone, di mana algoritma C4.5 dilatih menggunakan dataset yang telah melalui tahap preprocessing dan pembagian data. Pelatihan model bertujuan untuk membangun pohon keputusan yang mampu mengklasifikasikan kebutuhan pengguna smartphone dengan akurat berdasarkan atribut-atribut yang telah ditentukan. Proses pelatihan melibatkan serangkaian perhitungan matematis untuk menentukan atribut terbaik sebagai pemisah data pada setiap node, serta evaluasi performa model untuk memastikan kualitas klasifikasi yang dihasilkan.

3.1.5.1 Alur Kerja Algoritma C4.5

Proses klasifikasi kebutuhan pengguna smartphone menggunakan algoritma C4.5 dilakukan secara sistematis melalui serangkaian tahapan yang saling berkaitan dan terintegrasi.



Gambar 3.2 Alur Kerja Algoritma Decision Tree C4.5





Proses klasifikasi smartphone dengan algoritma C4.5 dimulai dari dataset 500 data dengan 4 kelas target, yang diproses melalui preprocessing dan feature engineering menjadi 14 fitur final, lalu dibagi secara stratified menjadi data latih 70%, validasi 15%, dan data uji 15%. Sebelum training, ditentukan hyperparameter fixed (criterion='entropy', splitter='best', class_weight='balanced', max_features=None) dan hyperparameter yang dituning via grid search (max_depth, min_samples_split, min_samples_leaf, ccp_alpha). Pembangunan pohon secara rekursif dilakukan menggunakan pustaka scikit-learn dengan kriteria Information Gain berbasis entropy (criterion='entropy'), sebagai pendekatan terhadap algoritma C4.5, mengingat pustaka tersebut tidak menyediakan Gain Ratio secara native pada proses pembentukan pohon. Proses berlanjut hingga stopping criteria terpenuhi dan terbentuk leaf node berisi distribusi kelas. Hasilnya berupa parameter otomatis: fitur split, threshold, struktur pohon, dan distribusi kelas tiap leaf.

Model dievaluasi menggunakan metrik dan cross validation, lalu digunakan untuk memprediksi kelas smartphone ke dalam empat kategori (Gaming, Fotografi/Videografi, Pelajar, Pengguna Umum), dan di-deploy sebagai web app. Entropy digunakan untuk mengukur tingkat ketidakpastian atau kekacauan dalam dataset. Semakin tinggi nilai entropy, semakin besar ketidakpastian data. Rumus entropy adalah:

$$Entropy(s) = -\sum p_i \times \log_2(p_i) \quad (1)$$

Dimana:

S = himpunan kasus/dataset

p_i = proporsi kelas i dalam dataset S

n = jumlah kelas

Information gain mengukur seberapa besar pengurangan entropy yang dihasilkan oleh suatu atribut ketika digunakan untuk memisahkan data. Atribut dengan information gain tertinggi dipilih sebagai node pemisah

Rumus Information Gain:

$$Gain(S, A) = Entropy(S) - \sum ((|S_v|/|S|) \times Entropy(S_v)) \quad (2)$$

Dimana:

S = himpunan kasus

A = atribut yang akan diuji

S_v = subset dari S untuk nilai v dari atribut A

$|S_v|$ = jumlah kasus dalam S_v

$|S|$ = jumlah total kasus dalam S

Gain ratio merupakan perbaikan dari information gain yang mengurangi bias terhadap atribut dengan banyak nilai. Pada penelitian ini, perhitungan Gain Ratio tidak dilakukan secara otomatis oleh pustaka scikit-learn, melainkan dihitung secara manual sebagai validasi terhadap pemilihan fitur pada node akar (root node).

Rumus gain ratio adalah

Rumus Gain Ratio:

$$Gain Ratio(S, A) = Gain(S, A) / Splitinfo(S, A) \quad (3)$$





1. Dilarang memperbanyak atau mendistribusikan dokumen ini untuk tujuan komersial tanpa izin tertulis dari penulis atau pihak berwenang.
2. Penggunaan tanpa izin untuk kepentingan komersial atau pelanggaran hak cipta dapat dikenakan sanksi sesuai dengan UU Hak Cipta di Indonesia.
3. Universitas hanya berhak menyimpan dan mendistribusikan dokumen ini di repositori akademik, tanpa mengalihkan hak cipta penulis, sesuai dengan peraturan yang berlaku di Indonesia.

Split info digunakan untuk mengukur potensi informasi yang dihasilkan dari pembagian dataset berdasarkan suatu atribut. Perhitungan Entropy dan Information Gain dilakukan secara rekursif pada setiap node oleh scikit-learn hingga memenuhi kriteria stopping, yaitu ketika semua data dalam node memiliki kelas yang sama, mencapai batas `max_depth`, atau tidak ada lagi pembagian yang memenuhi syarat `min_samples_split/min_samples_leaf`. Perhitungan Gain Ratio dan Split Info di atas digunakan secara terpisah, satu kali pada node akar, sebagai validasi kesesuaian hasil split otomatis dengan prinsip C4.5, sebagaimana ditunjukkan pada Tabel 4.4 Dimana Split Info adalah:

$$Split Info(S, A) = -\sum \left(\left(\frac{|S_v|}{|S|} \right) \times \log^2 \left(\frac{|S_v|}{|S|} \right) \right) \quad (4)$$

3.1.5.2 Evaluasi Model

Setelah model algoritma C4.5 selesai dilatih menggunakan data training, tahap selanjutnya adalah mengevaluasi performa model untuk memastikan akurasi dan keandalan sistem rekomendasi. Evaluasi model dilakukan menggunakan confusion matrix, yaitu matriks yang menampilkan perbandingan antara hasil prediksi model dengan nilai aktual dari data testing. Confusion matrix memberikan gambaran detail mengenai seberapa baik model dalam mengklasifikasikan setiap kategori kebutuhan pengguna smartphone.

1. Accuracy (Akurasi)

Akurasi mengukur proporsi prediksi yang benar dari keseluruhan prediksi yang dilakukan. Rumus akurasi adalah:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

2. Precision (Presisi)

Precision mengukur proporsi prediksi positif yang benar dari seluruh prediksi positif yang dilakukan model. Rumus precision adalah:

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

3. Recall (Sensitivitas)

Recall mengukur proporsi data positif yang berhasil diprediksi dengan benar oleh model dari seluruh data positif yang sebenarnya. Rumus recall adalah

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

4. F1-Score

F1-Score adalah rata-rata harmonis dari precision dan recall, yang memberikan keseimbangan antara kedua metrik tersebut. Rumus F1-Score adalah:

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (8)$$

3.1.6 Perancangan Sistem

Tahap perancangan menggambarkan alur kerja dan interaksi pengguna dengan sistem sebelum implementasi, menggunakan use case diagram, activity diagram Sequence diagram dan class diagram.



3.1.7 Implementasi Sistem

Implementasi merupakan realisasi hasil analisis dan perancangan menggunakan pendekatan Waterfall. Model C4.5 dilatih menggunakan dataset smartphone yang telah diproses, kemudian disimpan dalam bentuk file model (pickle) untuk proses prediksi. Sistem web dikembangkan menggunakan backend Flask yang menyediakan endpoint untuk memuat model dan memproses klasifikasi, serta frontend dengan desain responsive untuk menampilkan antarmuka kepada pengguna.

3.1.8 Pengujian Sistem

Pengujian dilakukan untuk mengevaluasi kinerja dan keakuratan sistem. Decision Tree C4.5 diuji menggunakan dataset testing dengan metrik accuracy, precision, recall, F1-score, dan confusion matrix untuk menganalisis distribusi serta kesalahan klasifikasi tiap kelas. Selain itu, dilakukan Cross Validation 5-Fold untuk mengevaluasi konsistensi performa model dan mendeteksi kemungkinan overfitting. Pengujian fungsional dilakukan dengan black-box testing untuk memastikan fitur input kriteria, proses klasifikasi, dan tampilan hasil rekomendasi berjalan dengan baik.

3.1.9 Kesimpulan Penelitian

Pada Tahap ini menyimpulkan hasil penelitian berdasarkan analisis data hingga pengujian sistem, mencakup evaluasi performa model C4.5 dalam mengklasifikasi kebutuhan pengguna, efektivitas sistem dalam membantu pelanggan Toko Sahabat Ponsel Tembilahan, serta saran pengembangan untuk penelitian selanjutnya.

