

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Perkembangan *smartphone* di Indonesia menunjukkan peningkatan yang sangat pesat, baik dari segi jumlah produk maupun variasi spesifikasi teknis yang ditawarkan berbeda. Setiap tahun produsen menghadirkan *smartphone* dengan perbedaan kapasitas ram, memori internal, resolusi kamera, ukuran layar, dan kapasitas baterai. Banyaknya varian spesifikasi tersebut menyebabkan proses pengelompokan *smartphone* ke dalam kategori segmen tertentu menjadi semakin kompleks apabila dilakukan secara manual (Artaningsih, 2022). Kondisi ini menuntut adanya pendekatan berbasis data yang mampu mengolah spesifikasi teknis secara terstruktur dan objektif.

Pengelompokan tersebut penting dilakukan karena dapat memberikan informasi mengenai posisi atau *segmen* suatu *smartphone* berdasarkan pola kombinasi spesifikasi yang dimiliki. Pengelompokan yang terstruktur dapat membantu proses identifikasi dan perbandingan produk, terutama ketika terdapat sejumlah *smartphone* dengan karakteristik spesifikasi yang saling tumpang tindih. Selain itu bertambahnya variasi produk menyebabkan pendekatan manual menjadi kurang efisien dan berpotensi menghasilkan penilaian yang tidak konsisten. Oleh karena itu, diperlukan pendekatan berbasis data yang mampu mempelajari pola hubungan antara spesifikasi teknis dan kelas *smartphone* sehingga proses klasifikasi terhadap data baru dapat dilakukan secara lebih sistematis dan objektif.

Salah satu pendekatan yang dapat digunakan untuk mengatasi kompleksitas tersebut adalah teknik klasifikasi dalam data *mining* dan *machine learning*. Klasifikasi merupakan proses pengelompokan data ke dalam kelas-kelas tertentu berdasarkan atribut yang dimiliki, dengan kelas target telah ditentukan sebelumnya. Pada konteks *smartphone*, teknik klasifikasi dapat dimanfaatkan untuk mengelompokkan ke dalam segmen seperti (*Entry*, *Mid*, dan *Flagship*) (Ramadhani dkk., 2025). Pendekatan ini memungkinkan data dengan jumlah dan atribut yang beragam untuk dianalisis secara sistematis sehingga pola antar spesifikasi dan kelas *smartphone* dapat dipelajari dengan

lebih akurat.

Penelitian terkait klasifikasi *smartphone* berbasis spesifikasi teknis telah dilakukan oleh Jarkhi (2025) dengan menggunakan algoritma *Naïve Bayes*. Penelitian tersebut memanfaatkan beberapa atribut spesifikasi untuk melakukan klasifikasi *smartphone* kedalam kategori harga tertentu. Hasil penelitian tersebut menunjukkan bahwa algoritma *Naïve Bayes* mampu melakukan klasifikasi berdasarkan pola yang terdapat pada atribut spesifikasi yang digunakan dengan tingkat akurasi yang diperoleh dalam penelitian tersebut. Temuan ini menunjukkan bahwa spesifikasi teknis *smartphone* dapat dimanfaatkan sebagai atribut dalam proses klasifikasi dan menjadi salah satu dasar bagi pengembangan penelitian klasifikasi *smartphone* menggunakan metode *machine learning* lainnya.

Meskipun penelitian Jarkhi (2025) menunjukkan bahwa algoritma *Naïve Bayes* dapat digunakan untuk mengklasifikasi *smartphone* berdasarkan atribut spesifikasi teknis, penelitian tersebut masih menggunakan satu pendekatan algoritma klasifikasi. *Naïve Bayes* bekerja berdasarkan asumsi independensi kondisional atribut terhadap kelas, sehingga karakteristiknya berbeda dengan algoritma berbasis pohon keputusan yang mampu membentuk aturan klasifikasi melalui proses pemisahan data berdasarkan atribut tertentu. Oleh karena itu, penelitian ini menggunakan algoritma *Decision Tree C4.5* dan *Random Forest* sebagai pendekatan alternatif dalam mengklasifikasikan *smartphone* berdasarkan spesifikasi teknis. *Decision Tree C4.5* digunakan karena menghasilkan struktur pohon keputusan yang dapat diinterpretasikan, sedangkan *Random Forest* menggunakan metode *ensembl learning* dengan membangun banyak pohon keputusan dan menggabungkan hasil prediksinya. Selain itu, dilakukan *hyperparameter tuning* pada *Random Forest* untuk mengetahui pengaruh pengaturan parameter terhadap kinerja model klasifikasi.

Penelitian mengenai penerapan algoritma berbasis pohon pada objek *smartphone* telah dilakukan pada beberapa penelitian sebelumnya. Diana dkk. (2025) menerapkan algoritma *Decision Tree C4.5* dalam sistem rekomendasi pembelian *smartphone* dengan memanfaatkan atribut seperti prosesor, kamera utama, kamera depan, ram, memori internal dan harga. Hasil penelitian tersebut menunjukkan bahwa *Decision*

Tree C4.5 dapat digunakan untuk membentuk aturan keputusan berdasarkan karakteristik spesifikasi *smartphone*.

Penerapan *Random Forest* pada objek yang serupa dilakukan oleh Aulia & Wahyuni (2024) untuk melakukan pemilihan dan klasifikasi *smartphone* berdasarkan atribut spesifikasi teknis, antara lain kamera, *chipset*, RAM, resolusi layar dan dukungan perekaman video. Penelitian tersebut menunjukkan bahwa *Random Forest* dapat digunakan untuk mengolah *sejumlah* atribut spesifikasi dalam proses klasifikasi perangkat *smartphone*. Selain itu, Liguori dkk. (2025) menerapkan *Random Forest* untuk klasifikasi rentang harga ponsel berdasarkan spesifikasi teknis ke dalam kategori *low*, *mid range*, dan *flagship*. Dengan menggunakan 2000 data, model yang dikembangkan memperoleh akurasi sebesar 89%, dengan ram sebagai salah satu atribut yang memiliki kontribusi dominan dalam proses klasifikasi. Kesamaan objek dan penggunaan atribut spesifikasi pada penelitian ini untuk menerapkan *Decision Tree C4.5* dan *Random Forest* dalam klasifikasi *smartphone* kedalam kelas *Entre-Level*, *Mid Range*, dan *Flagship* berdasarkan spesifikasi teknisnya.

Meskipun *Random Forest* telah diterapkan pada klasifikasi *smartphone* berdasarkan spesifikasi teknis, masih terdapat ruang pengembangan pada proses optimasi model. Beberapa penelitian terdahulu lebih fokus pada penerapan evaluasi algoritma, sementara eksplorasi terhadap konfigurasi *hyperparameter* belum menjadi fokus utama. Padahal pengaturan *hyperparameter* seperti jumlah pohon dan kedalaman maksimum pohon dapat memengaruhi kompleksitas serta kinerja model *Random Forest*. Oleh karena itu, penelitian ini menerapkan algoritma *Decision Tree C4.5* dan *Random Forest* untuk mengklasifikasi *smartphone* berdasarkan spesifikasi teknis, serta melakukan *hyperparameter tuning* pada *Random Forest* menggunakan *GridSearchCV*. Proses *tuning* dilakukan untuk mengevaluasi apakah perubahan konfigurasi parameter dapat memberikan perubahan terhadap kinerja klasifikasi dibandingkan model *Random Forest* sebelum *tuning*. Selain itu, penelitian ini juga menganalisis tingkat kepentingan fitur (*feature importance*) pada model *Random Forest* untuk mengetahui terhadap proses klasifikasi *smartphone*.

Pada implementasi penelitian ini, *Decision Tree C4.5* digunakan untuk

menganalisis atribut yang paling informatif melalui perhitungan *entropy*, *information gain*, *split information*, dan *gain ratio*. Sementara itu, *Random Forest* digunakan sebagai model klasifikasi utama melalui dua skenario, yaitu parameter *baseline* dan parameter hasil *hyperparameter tuning*. Pemisahan peran tersebut dilakukan agar penelitian tidak hanya memperoleh model klasifikasi, tetapi juga memperoleh interpretasi awal mengenai atribut spesifikasi yang paling berpengaruh dalam pembentukan keputusan.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang sudah dijelaskan sebelumnya, maka ada beberapa rumusan masalah pada penelitian kali ini adalah sebagai berikut :

1. Bagaimana cara mengimplementasikan algoritma *Decision Tree C4.5* dan *Random Forest* dalam mengklasifikasikan *smartphone* berdasarkan data spesifikasinya?
2. Bagaimana pengaruh *hyperparameter tuning* terhadap hasil klasifikasi *Smartphone* dengan *Random Forest* ?

1.3 Tujuan Penelitian

Adapun tujuan penelitian dari penelitian kali ini adalah sebagai berikut:

1. Menerapkan algoritma *Decision Tree C4.5* sebagai pemilihan *root node* dan *Random Forest* sebagai model dalam mengklasifikasikan *smartphone* berdasarkan spesifikasi teknisnya.
2. Menganalisis pengaruh *hyperparameter tuning* terhadap kinerja klasifikasi *Random Forest* dengan membandingkan hasil sebelum dan sesudah proses *hyperparameter tuning* berdasarkan nilai *Accuracy*, *Precision*, *Recall*, dan *F1-Score*

1.4 Manfaat Penelitian

Penelitian ini diharapkan memberi beberapa manfaat baik secara teoritis maupun praktis, antara lain:

1. Menambah referensi ilmiah di bidang *machine learning*, khususnya pada penerapan algoritma *Decision Tree C4.5* dan *Random Forest* dengan *hyperparameter tuning* pada permasalahan klasifikasi data.
2. Memberikan gambaran penerapan klasifikasi *smartphone* berdasarkan spesifikasi teknis secara objektif menggunakan pendekatan *supervised learning*.
3. Membantu memahami pengelompokan kelas *smartphone* (*Entry*, *Mid*, dan *Flagship*) berdasarkan karakteristik spesifikasinya.
4. Menjadi acuan bagi peneliti selanjutnya dalam pengembangan dan optimasi model klasifikasi berbasis data.

1.5 Batasan Penelitian

Ruang lingkup penelitian dibatasi pada hal-hal berikut:

1. Dataset yang digunakan berasal dari sumber publik seperti *kaggle*, yang berisi data spesifikasi *smartphone* dengan jumlah 3260. Dataset yang digunakan diambil di *kaggle* yang dibagikan oleh Roshan Yadav.
2. Klasifikasi ini dibangun menggunakan algoritma *Decision Tree C4.5* dan *Random Forest* dengan *Hyperparameter tuning*.
3. *Decision Tree C4.5* digunakan untuk analisis *root node* berdasarkan *Gain Ratio*, sedangkan model klasifikasi utama menggunakan model yang terbaik antara *Random Forest baseline* dan *Random Forest Tuning*.