

BAB I

PENDAHULUAN

1.1 Latar Belakang

Peran teknologi digital dalam pertumbuhan kebutuhan manusia modern saat ini semakin tinggi dan kompleks, hal tersebut memicu perusahaan-perusahaan digital agar terus melakukan pengembangan fitur pada produk yang mereka hasilkan. Salah satu teknologi digital yang saat ini cukup banyak diminati adalah *smartphone*. Hal ini dapat diketahui dari riset yang telah dilakukan TNS Australia dan Google yang menunjukkan 50% pengguna *smartphone* di Indonesia menggunakan teknologi digital tersebut sebagai alat telekomunikasi utama. Angka ini membuat Indonesia menduduki peringkat pertama dibanding dengan negara-negara di kawasan Asia Pasifik [1].

Tidak jauh berbeda dengan perangkat teknologi digital lain, *smartphone* juga terus mengalami perkembangan dalam hal fitur-fitur yang disediakan. Menurut Setijadi [2], perkembangan industri *smartphone* di Indonesia mengalami peningkatan dalam bidang fitur yang canggih. Hal ini memicu munculnya persaingan ketat dari pelaku usaha di bidang telekomunikasi, karena fitur produk sendiri berpengaruh positif dan dominan dalam keputusan untuk melakukan pembelian *smartphone*. Selain fitur yang diberikan, terdapat variabel harga yang memiliki pengaruh dalam keputusan untuk melakukan pembelian [3]. Oleh karena itu, para pelaku usaha baru di bidang industri *smartphone* membutuhkan klasifikasi untuk memprediksi harga *smartphone*, berdasarkan data harga *smartphone* yang saat ini telah tersedia.

Menurut Putra, dkk [4], Beragamnya fitur- fitur yang disediakan di dalam produk *smartphone* membantu untuk menentukan harga *smartphone* apakah termasuk dalam klasifikasi murah atau mahal. Metode klasifikasi sendiri adalah pendekatan pada fungsi klasifikasi yang dapat digunakan untuk melakukan prediksi atas suatu informasi yang belum diketahui dari sebelumnya. [5] Dalam klasifikasi, terdapat beberapa metode yang dapat digunakan, di antaranya adalah *Decision Tree*, *Naïve Bayes*, *K-Nearest Neighbors* (KNN), dan *Random Forest* [6]. Pada penelitian ini, metode klasifikasi yang digunakan adalah *Random Forest*. Pemilihan algoritma ini didasari oleh adanya beberapa penelitian terdahulu yang

membahas tentang perbandingan kinerja dari beberapa metode klasifikasi, yaitu penelitian yang dilakukan oleh Fajar Sodik Pamungkas, dkk [7], serta pada penelitian yang dilakukan oleh Intan Kemala, dkk[8]. Hasil kesimpulan dari kedua penelitian tersebut menyatakan bahwa algoritma *Random Forest* memiliki nilai akurasi tertinggi dibandingkan dengan algoritma lain yang diujikan.

Penelitian lain yang menggunakan algoritma *Random Forest* adalah penelitian yang dilakukan oleh Mulyana, dkk [6]. Penelitian ini memakai *dataset* dengan tujuh variabel prediksi. Kesimpulan yang didapat, menyatakan bahwa tingkat akurasi yang diperoleh dari penggunaan algoritma *Random Forest* adalah sebesar 81%. Penelitian tersebut tidak menggunakan seleksi fitur dalam pengolahan *dataset* yang dipakai.

Seleksi fitur sendiri merupakan proses untuk memilih sejumlah fitur yang paling signifikan atau berpengaruh terhadap akurasi dalam klasifikasi [9]. Tujuan dari penggunaan seleksi fitur adalah untuk mengurangi waktu komputasi dari metode klasifikasi yang digunakan, serta tanpa merusak nilai akurasi yang dihasilkan. Berdasarkan penelitian yang dilakukan oleh Anirma Kandida Br Ginting, dkk [10], menunjukkan adanya peningkatan nilai akurasi klasifikasi dari KNN, dengan peningkatan rata-rata sebesar 3.00 %, yang diperoleh setelah dilakukannya seleksi fitur pada *dataset*, sebelum akhirnya digunakan dalam proses klasifikasi. Pada penelitian yang dilakukan oleh Divinda Morikawa [11], hasil pengujian yang didapat menunjukkan bahwa penggunaan metode *forward selection* mampu menghasilkan peningkatan kinerja yang lebih tinggi dibandingkan dengan metode seleksi fitur lain yang diujikan.

Berdasarkan analisa pada hasil penelitian terdahulu, penulis akan melakukan penelitian yang berjudul “Analisis Penggunaan Seleksi Fitur *Forward Selection* Untuk Masalah Waktu Komputasi Pada Metode *Random Forest* (Studi Kasus: Klasifikasi Harga *Smartphone*)”. Penelitian ini bertujuan untuk menilai kinerja metode seleksi fitur *forward selection* dalam mengurangi waktu komputasi dari metode *random forest* pada klasifikasi harga *smartphone*.

1.2 Rumusan Masalah

Adapun rumusan masalah yang diangkat dalam penelitian ini dibagi menjadi tiga sub-bab, sebagai berikut:

1.2.1 Permasalahan

Semakin banyak jumlah fitur pada *dataset*, semakin lama pula waktu komputasi dari sistem klasifikasi yang dijalankan. Oleh karena itu diperlukan metode untuk mengurangi jumlah fitur yang digunakan pada proses klasifikasi, tetapi tidak secara signifikan berpengaruh terhadap nilai akurasi yang dihasilkan. Hal ini dapat dilakukan dengan memilih hanya fitur-fitur yang memiliki pengaruh lebih tinggi terhadap label kelas, sehingga proses komputasi yang berlangsung menjadi lebih cepat.

1.2.2 Solusi Permasalahan

Penerapan metode seleksi fitur *forward selection* untuk mendapatkan fitur-fitur dengan nilai pengaruh tinggi, sehingga dapat mengurangi waktu komputasi yang diperlukan, tanpa secara signifikan mempengaruhi nilai akurasi yang dihasilkan.

1.2.3 Pertanyaan Penelitian

1. Bagaimana perbandingan waktu komputasi dari sistem klasifikasi *random forest* dalam mengklasifikasikan harga *smartphone*, sebelum dan sesudah penggunaan seleksi fitur *forward selection*?
2. Bagaimana perbandingan nilai akurasi yang dihasilkan, sebelum dan sesudah penggunaan seleksi fitur *forward selection*?

1.3 Manfaat dan Tujuan

Adapun manfaat dan tujuan penelitian ini sebagai berikut:

1.3.1 Tujuan

Mengetahui kinerja metode seleksi fitur *forward selection* dalam mengurangi waktu komputasi dari uji klasifikasi *random forest* dalam mengklasifikasikan harga *smartphone*, serta mengetahui perbandingan hasil akurasi metode *random forest* antara sebelum dan sesudah menggunakan seleksi fitur *forward selection*.

1.3.2 Manfaat

Diharapkan mampu memudahkan pelaku usaha dibidang telekomunikasi khususnya *smartphone* dalam memprediksi harga, berdasarkan pertimbangan terhadap fitur-fitur yang mempengaruhi harga *smartphone*.

1.4 Batasan Masalah

1. *Dataset* yang digunakan adalah data klasifikasi harga ponsel yang telah disediakan di <https://www.kaggle.com/datasets/iabhishekofficial/mobile-price-classification>.
2. *Dataset* yang digunakan memiliki 401 data dan 18 fitur.

1.5 Sistematika Laporan

Adapun sistematika penyusunan laporan skripsi ini sebagai berikut:

BAB I PENDAHULUAN

Bab ini membahas mengenai latar belakang permasalahan, rumusan masalah, manfaat, tujuan yang ingin dicapai, metode yang digunakan, batasan masalah dalam penyusunan skripsi, serta sistematika penulisan laporan.

BAB II KAJIAN PUSTAKA

Bab ini menjelaskan mengenai teori dari penelitian terdahulu dan *literature* yang dapat menunjang dalam penyusunan penelitian.

BAB III METODE USULAN

Bab ini menyajikan jenis *dataset*, metode, dan langkah-langkah dalam penyusunan penelitian yang dilakukan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi pembahasan tentang langkah-langkah pengolahan data dan analisa hasil uji coba pada sistem yang telah dibuat.

BAB V PENUTUP

Bab ini berisi kesimpulan dari hasil analisa yang telah dilakukan, serta saran pengembangan untuk penelitian selanjutnya.