

BAB I

PENDAHULUAN

1.1. LATAR BELAKANG MASALAH

Data mining merupakan proses mengekstraksi pengetahuan atau informasi yang berguna dari kumpulan data yang besar dan kompleks [1]. *Data mining* tidak hanya mengumpulkan data tetapi juga mencakup analisis dan analisis prediktif terhadap informasi yang ingin ditampilkan. Data yang dikumpulkan akan disimpan dalam database untuk menemukan keteraturan, pola baru yang valid atau hubungan dalam data dan dapat dipahami dalam database besar untuk pengambilan keputusan ketika mempertimbangkan informasi apa yang akan digunakan [2][3]. Dimana proses metode *data mining* dapat mendeskripsikan temuan baru dalam suatu kumpulan data dan menjelaskan proses penggunaan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstrak dan mengidentifikasi informasi bermanfaat serta mempunyai fungsi diantaranya *Clustering*, *Association Sequencing*, *Forecasting*, dan *Classification* [4][5]. Selain itu, dalam proses *data mining* ini menggunakan metode klasifikasi.

Klasifikasi adalah proses menemukan sekumpulan pola atau fungsi yang mendeskripsikan dan memisahkan suatu kelas data dengan kelas data lainnya dan digunakan untuk memprediksi kelas suatu objek yang labelnya tidak diketahui [6], [7]. Klasifikasi termasuk dalam metode pembelajaran atau *supervised* karena memerlukan pembelajaran dari data sebelumnya untuk mengetahui hasil dari data baru [8]. Dalam proses ini, algoritma pembelajaran mengidentifikasi pola yang

menghubungkan atribut data dengan kelas yang sesuai, sehingga memungkinkan untuk membuat prediksi berdasarkan pola-pola ini.

Supervised Learning adalah Algoritma pembelajaran yang diawasi membangun model matematika dari kumpulan data yang berisi masukan dan keluaran yang diinginkan [9]. Terdapat banyak metode yang ada dalam klasifikasi *supervised learning* diantaranya adalah *K-Nearest Neighbors* (K-NN) [10]. Seiring dengan kemajuan teknologi dan peningkatan dalam penggunaan data, semakin banyak metode pembelajaran *supervised* seperti K-NN digunakan untuk mengatasi berbagai tantangan analisis data.

K-Nearest Neighbor (K-NN) merupakan metode *supervised learning* yang artinya memerlukan informasi training untuk mengklasifikasikan objek yang jaraknya sangat dekat [11]. Prinsip pengoperasian K-NN adalah menemukan k tetangga terdekat dan mengklasifikasikan rekaman data *numerik* dengan menghitung jarak antara sampel uji dan seluruh sampel pelatihan menggunakan jarak *Euclidian* yang dapat digunakan pada *dataset* harga ponsel (*mobile price*) [12]. K-NN adalah salah satu metode yang sederhana tetapi efektif dalam klasifikasi, dan sering digunakan dalam berbagai aplikasi, termasuk klasifikasi pemrosesan gambar, produk, dan rekomendasi.

Dataset mobile price didapat dari *website kaggle.com* mempunyai 2000 data, 20 atribut serta 1 kelas yaitu *price range*. Ponsel atau telepon adalah perangkat komunikasi (seperti menelepon atau mengirim SMS) yang juga mencakup fungsi PDA (*Personal Digital Assistant*) dan mungkin menyerupai

komputer. Kenyamanan yang dimiliki ponsel ini menjadikannya sebagai alat komunikasi yang dibutuhkan setiap orang [13]. Saat ini ponsel memiliki banyak fitur yang berbeda-beda, seiring berkembangnya teknologi, ponsel memiliki lebih banyak fungsi dan spesifikasi yang lebih baik. Ciri-ciri ponsel seperti resolusi kamera, bentuk ramping, layar lebar, kapasitas memori besar, dll dapat menentukan harga ponsel [14]. Analisis data pada dataset ini dapat membantu produsen ponsel untuk memahami tren dan preferensi konsumen dalam membeli ponsel serta merancang produk yang sesuai dengan kebutuhan pasar.

Harga adalah tolak ukur pertama saat membeli atau menjual. Pelanggan yang ingin membeli telepon seluler pasti mempertimbangkan harga telepon seluler yang akan dibelinya berdasarkan spesifikasinya. Oleh karena itu, pengklasifikasian harga ponsel sangat diperlukan sebagai referensi pelanggan ponsel dalam menentukan apakah ponsel tersebut sesuai dengan keinginan pelanggan atau tidak. Keuntungan lain dalam melakukan pengklasifikasian harga ponsel yaitu pelanggan dapat memilah apakah ponsel tersebut layak digunakan pada kemajuan teknologi saat ini mengingat beragam macam spesifikasi ponsel yang ada pada saat ini [14]. Dengan pemahaman yang lebih baik tentang faktor-faktor yang memengaruhi harga ponsel, pelanggan dapat membuat keputusan yang lebih cerdas dalam memilih ponsel yang sesuai dengan kebutuhan dan anggaran mereka, sementara produsen dapat mengarahkan upaya mereka untuk mengembangkan produk yang lebih kompetitif di pasar yang berubah dengan cepat.

Menurut Aditya Arisusanto et al [15] dalam memprediksi *mobile price* dengan algoritma *random forest* dengan *Optimize Parameter Grid* berdasarkan 2000 record dan terdiri dari 21 atribut mendapatkan hasil nilai *accuracy* sebesar 88,50% diperoleh dari nilai parameter *maximal_depth optimal* sebesar 12 dan 25, serta parameter *criterion* yang optimal pada *gini_index* pada klasifikasi data harga handphone menggunakan Algoritma *Random Forest* dengan *Optimize Parameter Grid*.

Sehubungan dengan latar belakang diatas, maka penulis tertarik untuk melakukan penelitian untuk mendapatkan akurasi yang lebih baik lagi dari penelitian sebelumnya dengan berjudul **“Implementasi *Data Mining* Dengan Algoritma K-Nearest Neighbors Untuk Klasifikasi Harga Ponsel”**.

1.2. RUMUSAN MASALAH

Berdasarkan dari latar belakang yang telah diuraikan di atas, maka di dapat rumusan masalah sebagai berikut :

1. Bagaimana penerapan *data mining* dengan algoritma *K-Nearest Neighbor* (K-NN) dalam mengklasifikasi *mobile price*?
2. Seberapa besar tingkat akurasi algoritma *K-Nearest Neighbor* (K-NN) apabila digunakan untuk memprediksi kelas *mobile price*?

1.3. BATASAN MASALAH

Untuk menghindari permasalahan yang meluas, maka penulis membatasi pembahasan permasalahan sebagai berikut :

1. Data yang digunakan bersumber dari *repository public* yaitu *kaggle.com* dengan nama *dataset mobile price classification*.
2. Metode yang digunakan adalah metode klasifikasi serta algoritma yang digunakan yaitu *K-Nearest Neighbor* (K-NN).
3. Kinerja algoritma *K-Nearest Neighbor* (K-NN) diukur dengan menggunakan nilai *accuracy*, *precision*, dan *recall*.
4. Aplikasi yang digunakan adalah *RapidMiner Studio*.

1.4. TUJUAN DAN MANFAAT PENELITIAN

1.4.1. Tujuan Penelitian

Adapun tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut :

1. Untuk menerapkan *data mining* dengan algoritma *K-Nearest Neighbor* (K-NN) dalam mengklasifikasi *mobile price*.
2. Untuk mengetahui seberapa akurat tingkat akurasi algoritma *K-Nearest Neighbor* (K-NN) dalam memprediksi kelas *mobile price*.

1.4.2. Manfaat Penelitian

Adapun beberapa manfaat yang akan didapat dalam melakukan penelitian ini, yaitu :

1. Dapat membantu mengetahui atribut yang mempunyai korelasi baik dan tidak pada *dataset mobile price*.
2. Dapat mengetahui performa dari algoritma *K-Nearest Neighbor* (K-NN) dalam memprediksi kelas *mobile price*.
3. Memberikan kontribusi keilmuan pada penelitian bidang klasifikasi *data mining* khususnya untuk klasifikasi *mobile price*.
4. Dapat digunakan sebagai acuan untuk melakukan penelitian berikutnya.

1.5. SISTEMATIKA PENULISAN

Sistematika penulisan menggambarkan secara umum tentang apa yang akan dibahas dalam setiap bab dari laporan penelitian ilmiah ini. Adapun isi pokok bahasan masing-masing bab sebagai berikut :

BAB I : PENDAHULUAN

Pada bab ini berisi tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan yang digunakan dalam penulisan laporan tugas akhir ini.

BAB II : LANDASAN TEORI

Pada bab ini menjelaskan teori-teori yang mendasari pembahasan laporan secara khusus berisi definisi- definisi yang di dapat melalui studi pustaka seperti kutipan buku, jurnal dan lain-lain sebagai dasar dalam pembuatan laporan. Pada bab ini juga memuat tinjauan pustaka yang berisi penelitian-penelitian sebelumnya yang berhubungan dengan penelitian ini.

BAB III : METODOLOGI PENELITIAN

Pada bab ini menjelaskan tentang penjelasan dari metode-metode penelitian yang digunakan sebagai solusi penyelesaian masalah yang diangkat dalam penelitian.

BAB IV : ANALISIS

Pada bab ini menjelaskan tentang gambaran umum pemilihan *dataset* pada *kaggle.com*. Kemudian menerapkan algoritma *K-Nearest Neighbor* (K-NN) untuk mengetahui tingkat akurasi data menggunakan aplikasi *RapidMiner Studio*.

BAB V : KESIMPULAN

Pada bab ini berisikan kesimpulan yang diambil dari hasil analisis serta saran-saran yang mencakup keseluruhan dari hasil penelitian.