BAB II
LANDASAN TEORI

2.1. Sistem Informasi

Sistem informasi merupakan suatu sistem terintegrasi yang berfungsi untuk mengumpulkan, mengolah, dan menyebarkan informasi yang relevan untuk mendukung pengambilan keputusan dalam organisasi [1]. Sistem informasi yang dirancang untuk mengelola data harga, fitur, dan merek ponsel memiliki berbagai fungsi penting yang mendukung kebutuhan pengguna dan bisnis [2]. Secara umum, sistem ini berfungsi untuk mengumpulkan, menyimpan, memproses, dan menyajikan data sehingga dapat digunakan untuk pengambilan keputusan yang lebih baik sistem informasi [3]. Dalam konteks ponsel, data seperti harga, dan merek memiliki nilai strategis yang besar, baik bagi konsumen yang ingin membeli. Salah satu fungsi utama sistem ini adalah membantu pengguna dalam membandingkan ponsel berdasarkan kriteria tertentu. Konsumen sering kali bingung ketika harus memilih perangkat yang sesuai dengan kebutuhan mereka di tengah banyaknya pilihan [4]. Dengan sistem informasi, data harga, fitur seperti ukuran layar, kapasitas baterai, kualitas kamera, dan nama merek dapat disusun secara terstruktur. Konsumen dapat dengan mudah memfilter perangkat berdasarkan anggaran, merek favorit, atau fitur yang diinginkan. Hal ini tidak hanya menghemat waktu tetapi juga memastikan bahwa pilihan yang dibuat lebih tepat. Sebagai contoh, jika data menunjukkan bahwa konsumen cenderung memilih ponsel dengan kapasitas baterai besar di kisaran harga tertentu, produsen dapat

fokus merancang perangkat yang memenuhi preferensi tersebut. Bagi penjual, sistem informasi ini memungkinkan pengelolaan inventaris yang lebih efisien. Dengan data harga yang selalu diperbarui, penjual dapat menyesuaikan strategi pemasaran agar tetap kompetitif. Sistem ini juga Mempermudah pelacakan stok, sehingga penjual dapat mengetahui kapan harus melakukan pengisian ulang barang. Ketika data terintegrasi dengan baik, risiko kehabisan stok atau kelebihan stok dapat diminimalkan, yang pada akhirnya meningkatkan efisiensi operasional.

Lebih jauh lagi, sistem informasi ini dapat memberikan manfaat dalam hal transparansi dan kepercayaan konsumen [5]. Dengan menampilkan data harga yang akurat dan fitur produk yang jelas, konsumen merasa lebih yakin saat bertransaksi. Transparansi ini juga membantu mencegah terjadinya miskomunikasi atau klaim palsu terkait spesifikasi produk. Hal ini sangat penting, terutama dalam transaksi online, di mana konsumen tidak bisa langsung memeriksa produk secara fisik. Selain untuk kepentingan konsumen dan penjual, sistem ini juga bisa digunakan oleh analis industri atau pihak ketiga yang ingin memahami dinamika pasar ponsel secara lebih mendalam.

Salah satu contoh konkret adalah algoritma rekomendasi. Sistem informasi di platform ini mengumpulkan data pengguna, seperti pencarian produk, klik pada halaman tertentu, atau pembelian sebelumnya [6]. Dari data tersebut, algoritma menggunakan analisis machine learning untuk memprediksi ponsel mana yang kemungkinan besar diminati oleh pengguna tertentu. Sebagai ilustrasi, jika seorang pengguna sering mencari ponsel dengan fitur kamera berkualitas tinggi di kisaran harga tertentu, sistem akan merekomendasikan ponsel yang sesuai dengan kriteria

tersebut. Data fitur kamera, merek, dan harga diproses oleh algoritma untuk menyaring pilihan terbaik. Dengan begitu, pengalaman belanja menjadi lebih personal dan relevan bagi konsumen. Selain rekomendasi, sistem informasi juga digunakan untuk mendukung penentuan harga dinamis. Dalam pasar yang sangat kompetitif, harga ponsel dapat berubah dengan cepat. Dengan menganalisis data dari berbagai sumber, seperti harga pesaing, tingkat permintaan, dan stok barang, algoritma dapat menentukan harga optimal secara real-time.

Jika data menunjukkan bahwa ponsel dengan layar besar dan baterai tahan lama menjadi pilihan utama di segmen harga menengah, produsen atau distributor dapat menggunakan informasi ini untuk menyesuaikan strategi pemasaran atau mengarahkan pengembangan produk. Informasi seperti ini dihasilkan dari pengolahan algoritma yang menganalisis ribuan data transaksi dan ulasan pelanggan [7]. Di sisi lain, konsumen juga mendapatkan manfaat besar dari sistem ini melalui fitur pembanding produk. Misalnya, seorang pembeli dapat memilih tiga ponsel dari merek yang berbeda untuk dibandingkan. Sistem informasi memanfaatkan algoritma untuk menampilkan tabel perbandingan, yang mencakup data harga, fitur utama seperti resolusi kamera, kapasitas penyimpanan, prosesor, hingga ulasan pengguna [8]. Menurut Zuhra, sistem informasi mencakup pengolahan data transaksi harian yang mendukung operasi dan kegiatan strategis. Syaifunazhirin menekankan bahwa sistem informasi teknologi berperan penting dalam mengintegrasikan berbagai sistem untuk mencapai tujuan yang lebih besar. Selain itu, sistem informasi juga berfungsi untuk meningkatkan kinerja organisasi, seperti yang ditunjukkan dalam penelitian Safira dan Ratnawati mengenai sistem

informasi akuntansi yang meningkatkan kinerja bisnis. Lebih lanjut, sistem informasi dapat membantu dalam pengelolaan data dan informasi yang efisien, seperti yang diungkapkan oleh Hadiluwarso. yang menyoroti pentingnya sistem informasi dalam pengendalian internal organisasi [9]. Dengan demikian, sistem informasi tidak hanya berfungsi sebagai alat pengolahan data, tetapi juga sebagai sarana strategis untuk meningkatkan efektivitas dan efisiensi operasional dalam berbagai konteks organisasi. Sistem informasi dapat didefinisikan sebagai sistem yang mengadopsi pengetahuan manusia ke dalam komputer untuk menyelesaikan masalah dengan cara yang mirip dengan pakar. Sebagai contoh, sistem pakar dapat digunakan untuk mendiagnosa penyakit, seperti yang dijelaskan dalam penelitian tentang sistem pakar untuk penyakit aritmia Hanifa [10].

2.2. Knowledge Discovery in Database

Knowledge Discovery in Databases (KDD) adalah proses yang kompleks untuk menemukan pengetahuan yang berguna dari data yang tersimpan dalam basis data [11]. Knowledge Discovery in Database (KDD) proses yang digunakan untuk menemukan pola, informasi, atau wawasan yang berguna dari kumpulan data yang besar [12]. Dalam konteks minat konsumen, KDD berfungsi untuk memahami preferensi, kebutuhan, dan perilaku pembelian mereka berdasarkan data yang telah dikumpulkan. Proses ini biasanya melibatkan beberapa tahap, seperti pengumpulan data, pembersihan data, analisis, dan interpretasi hasil. Misalnya, sebuah ponsel memiliki jutaan data transaksi pelanggan, seperti produk yang sering dibeli, waktu pembelian, atau ulasan produk. Dengan menggunakan KDD, data ini dapat diolah untuk mengidentifikasi tren tertentu [13].

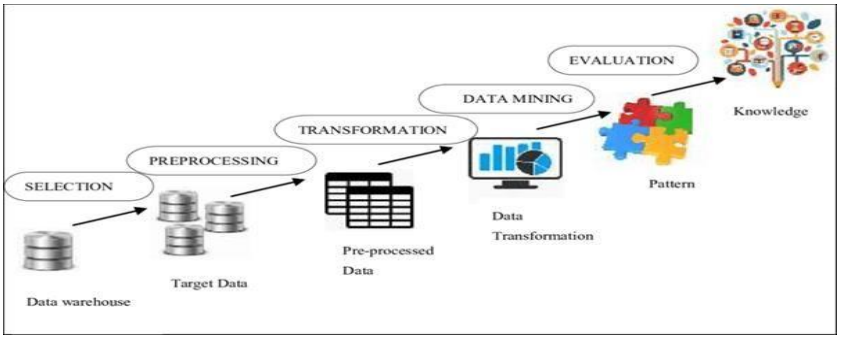
Contohnya, ponsel dapat menemukan bahwa konsumen cenderung membeli ponsel dengan fitur kamera berkualitas tinggi pada musim liburan atau bahwa pelanggan setia lebih tertarik pada penawaran.

Ketika kita ingin mengolah data tentang harga, merek, dan fitur ponsel, langkah-langkah dalam Knowledge Discovery in Databases (KKD) sangat relevan untuk mendapatkan informasi yang bermakna dari data tersebut. KKD adalah proses sistematis yang melibatkan beberapa tahapan, seperti pembersihan data, transformasi data, dan analisis, untuk menghasilkan wawasan yang berguna. Berikut penjelasannya:

Tahap pertama adalah pembersihan data (data cleaning). Dalam konteks data ponsel, data yang kita miliki mungkin tidak sempurna mungkin ada nilai yang hilang, kesalahan penulisan pada nama merek, atau data harga yang tidak masuk akal. Misalnya, jika ada harga ponsel yang tercatat sebagai "Rp 0" atau merek tertulis "Samung" alih-alih "Samsung," maka data tersebut perlu diperbaiki atau dihapus untuk memastikan keakuratannya. Selanjutnya adalah transformasi data (data transformation). Ini melibatkan pengubahan data mentah menjadi format yang lebih cocok untuk analisis. Sebagai contoh, harga ponsel yang awalnya tercatat dalam berbagai mata uang mungkin perlu dikonversi ke mata uang yang seragam. Selain itu, fitur-fitur ponsel yang tertulis dalam bentuk teks, seperti "RAM 8GB, Storage 128GB," bisa dipecah menjadi kolom terpisah untuk RAM dan penyimpanan agar lebih mudah dianalisis. Setelah data dibersihkan dan ditransformasi, kita dapat melanjutkan ke tahap-tahap analisis lainnya, seperti pengelompokan ponsel berdasarkan kisaran harga atau membandingkan popularitas

merek tertentu berdasarkan fitur unggulan. Langkah-langkah ini membantu kita memahami tren di pasar ponsel, seperti merek mana yang menawarkan fitur tertentu dalam kisaran harga tertentu atau bagaimana fitur-fitur tertentu memengaruhi harga ponsel.

Dengan menerapkan langkah-langkah KKD, data yang awalnya tidak terstruktur dan mungkin penuh dengan kesalahan dapat diubah menjadi sumber wawasan yang mendalam tentang pasar ponsel. Hasil dari proses KDD ini membantu perusahaan membuat strategi yang lebih efektif, seperti menyesuaikan promosi, meningkatkan personalisasi layanan, atau merancang produk yang sesuai dengan preferensi pasar. Dalam dunia bisnis yang sangat kompetitif, kemampuan untuk menemukan pola dari data besar melalui KDD menjadi keunggulan penting untuk memahami dan memenuhi kebutuhan konsumen secara lebih baik. Menurut Važan et al., KDD mencakup serangkaian langkah yang meliputi pembersihan data, integrasi data, pemilihan data, transformasi data, dan data mining sebagai langkah kunci dalam proses ini. Kumar dan Rathee menekankan bahwa KDD melibatkan ekstraksi informasi yang tidak jelas, sebelumnya tidak diketahui, dan berpotensi berguna dari data. Lebih lanjut, Amorim et al. mendefinisikan KDD sebagai proses penemuan pengetahuan yang valid, baru, dan dapat dipahami dari basis data. Fayyad et al. juga memberikan definisi yang umum diterima, yaitu proses nontrivial untuk mengidentifikasi pola yang valid, baru, dan berguna dalam data. Dengan demikian, KDD berfungsi sebagai alat penting dalam berbagai bidang, termasuk bisnis dan penelitian, untuk meningkatkan pengambilan keputusan berdasarkan data yang ada.



Gambar 2. 1. Proses KDD

Tahapan KDD:

Berikut adalah tahapan-tahapan dalam proses Knowledge Discovery in Databases (KDD) beserta penjelasannya masing-masing:

1. Selection (Seleksi Data): Tahapan ini bertujuan untuk memilih data yang relevan dari sumber data yang tersedia untuk dianalisis. Data yang dipilih harus sesuai dengan tujuan analisis dan memiliki atribut yang mendukung proses penggalian pengetahuan.
2. Preprocessing (Pra-pemrosesan): Pada tahap ini, data yang telah dipilih dibersihkan dari nilai-nilai yang tidak valid, hilang, atau tidak konsisten. Tujuannya adalah untuk memastikan kualitas data agar tidak memengaruhi akurasi hasil analisis.
3. Transformation (Transformasi Data): Data yang telah dibersihkan kemudian diubah ke dalam format atau struktur yang sesuai untuk proses data mining. Proses ini bisa melibatkan normalisasi, pengkodean, atau pembuatan atribut baru.
4. Data Mining (Penggalian Data): Tahapan inti dalam KDD ini menggunakan algoritma tertentu untuk menemukan pola, hubungan, atau informasi

penting dari data. Teknik yang digunakan bisa berupa klasifikasi, klastering, asosiasi, atau regresi, tergantung pada tujuan analisis.

5. Interpretation/Evaluation (Interpretasi dan Evaluasi): Hasil dari data mining kemudian dievaluasi untuk menilai validitas dan maknanya. Pola yang ditemukan dianalisis apakah benar-benar bermanfaat dan dapat dijadikan dasar pengambilan keputusan.

2.2.1. Klasifikasi

Klasifikasi dalam penelitian ini mengacu pada proses pengelompokan data pelanggan berdasarkan karakteristik tertentu yang dapat membantu dalam memahami faktor yang mempengaruhi keputusan pembelian mereka [14]. Dengan menerapkan algoritma Naïve Bayes, penelitian ini bertujuan untuk membangun model yang dapat mengkategorikan minat konsumen terhadap suatu merek ponsel berdasarkan berbagai fitur yang ditawarkan oleh ponsel tersebut [15]. Naïve Bayes merupakan algoritma klasifikasi berbasis probabilitas yang bekerja dengan prinsip Teorema Bayes, di mana setiap fitur dianggap independen satu sama lain dalam mempengaruhi keputusan akhir [16].

Dalam penelitian ini, klasifikasi berperan penting dalam memecahkan masalah bagaimana pola minat masyarakat Labuhanbatu terhadap merek ponsel tertentu dapat dipahami berdasarkan fitur yang paling mereka prioritaskan [17]. Dengan menerapkan metode ini, penelitian dapat menghasilkan wawasan mengenai faktor utama yang mendorong seseorang memilih suatu merek dibandingkan merek lainnya [18]. Proses klasifikasi dilakukan dengan mengumpulkan data dari berbagai konsumen, mengidentifikasi faktor-faktor utama dalam pengambilan keputusan

mereka, serta membangun model yang dapat mengelompokkan preferensi konsumen secara sistematis.

Penelitian ini membahas bagaimana fitur-fitur ponsel, seperti kapasitas baterai, kualitas kamera, kapasitas penyimpanan, harga, desain, dan performa mempengaruhi preferensi konsumen dalam memilih suatu merek [19]. Misalnya, ada segmen masyarakat yang lebih tertarik pada ponsel dengan kamera berkualitas tinggi, sementara yang lain lebih mengutamakan daya tahan baterai yang lama atau harga yang terjangkau. Dengan menggunakan klasifikasi, skripsi ini dapat mengelompokkan pola-pola yang terbentuk berdasarkan atribut-atribut tersebut dan memahami bagaimana konsumen menentukan pilihan mereka.

Proses klasifikasi dalam penelitian ini dimulai dengan pengumpulan data dari masyarakat Labuhanbatu melalui survei atau data transaksi yang mencerminkan preferensi konsumen dalam memilih merek HP. Data yang dikumpulkan kemudian diproses untuk memastikan bahwa informasi yang diperoleh bersih, tidak mengandung duplikasi, serta relevan dengan tujuan penelitian [20]. Setelah itu, data dipisahkan menjadi dua bagian, yaitu data latih (training data) yang digunakan untuk membangun model klasifikasi dan data uji (testing data) yang digunakan untuk mengukur seberapa akurat model dalam mengklasifikasikan preferensi konsumen. Penelitian ini menerapkan algoritma Naïve Bayes karena algoritma ini memiliki keunggulan dalam menangani masalah klasifikasi dengan jumlah data yang besar dan variabel independen yang banyak. Naïve Bayes sangat cocok untuk klasifikasi dalam skripsi ini karena dapat menangani data dengan banyak kategori, seperti berbagai merek HP dan fitur yang berbeda-beda. Algoritma ini bekerja

dengan menghitung probabilitas dari setiap kelas berdasarkan atribut yang dimiliki oleh data, sehingga mampu memberikan hasil klasifikasi yang cepat dan efisien.

Dalam penelitian ini, fitur-fitur ponsel yang menjadi fokus utama dalam klasifikasi dapat berupa spesifikasi teknis, harga, dan aspek lainnya yang mempengaruhi minat masyarakat. Dengan menggunakan Naïve Bayes, skripsi ini dapat memprediksi kecenderungan seseorang dalam memilih merek ponsel tertentu berdasarkan pola pembelian dan preferensi mereka sebelumnya. Misalnya, jika seseorang lebih sering membeli ponsel dengan baterai berkapasitas besar dan harga terjangkau, maka model yang dibangun dalam penelitian ini dapat mengklasifikasikan preferensi mereka dan memprediksi bahwa mereka akan memilih merek yang menawarkan fitur tersebut.

Dengan adanya model klasifikasi yang akurat, perusahaan dapat menyusun strategi pemasaran yang lebih efektif, seperti menargetkan iklan kepada segmen masyarakat yang lebih cenderung membeli produk tertentu atau menyesuaikan fitur ponsel dengan kebutuhan pasar. Selain itu, hasil dari penelitian ini dapat digunakan untuk meningkatkan keMinatan pelanggan dengan memberikan rekomendasi produk yang sesuai dengan preferensi mereka.

Dalam proses klasifikasi ini, terdapat beberapa tantangan yang dihadapi, salah satunya adalah memastikan bahwa data yang digunakan cukup representatif untuk menggambarkan pola minat konsumen di Labuhanbatu secara keseluruhan. Jika data yang dikumpulkan tidak mencerminkan populasi yang sebenarnya, maka hasil klasifikasi dapat menjadi bias atau tidak akurat.Selain itu, ini juga dijelaskan bagaimana metode klasifikasi dengan Naïve Bayes dibandingkan dengan metode

lain, seperti Decision Tree atau K-Nearest Neighbors (KNN). Naïve Bayes memiliki keunggulan dalam kecepatan dan kemudahan implementasi, namun memiliki keterbatasan dalam menangani data yang memiliki hubungan kompleks antar variabel. Oleh karena itu, penelitian ini juga membahas kemungkinan penggunaan metode hybrid atau kombinasi dengan teknik lain untuk meningkatkan akurasi klasifikasi. ini memberikan wawasan mendalam mengenai bagaimana klasifikasi dapat diterapkan dalam menganalisis pola minat konsumen terhadap pemilihan merek HP, terutama dalam konteks masyarakat Labuhanbatu. Dengan menggunakan pendekatan berbasis data, penelitian ini membantu mengidentifikasi faktor-faktor utama yang mempengaruhi keputusan pembelian konsumen serta memberikan rekomendasi yang lebih akurat berdasarkan pola yang ditemukan.

Secara keseluruhan, ini berkontribusi pada bidang ilmu data mining dan analisis konsumen dengan menunjukkan bagaimana klasifikasi dapat digunakan untuk memahami preferensi masyarakat terhadap teknologi. Dengan semakin berkembangnya industri ponsel dan meningkatnya persaingan antar merek, hasil dari penelitian ini dapat menjadi acuan bagi perusahaan untuk lebih memahami kebutuhan pelanggan dan menyesuaikan strategi bisnis mereka agar lebih sesuai dengan preferensi pasar.

Metode klasifikasi berbasis Naïve Bayes dapat membantu dalam mengembangkan model prediksi yang lebih efektif dalam memahami pola minat konsumen. Dengan demikian, hasil penelitian ini tidak hanya memberikan manfaat akademik, tetapi juga dapat diterapkan dalam dunia industri untuk mendukung pengambilan keputusan berbasis data yang lebih akurat dan efisien.

2.3. Metode *Naïve Bayes*

Algoritma Naive Bayes adalah metode klasifikasi yang berbasis pada teorema Bayes dengan asumsi independensi antar fitur [21]. Algoritma Naive Bayes adalah metode pembelajaran mesin yang digunakan untuk membuat prediksi berdasarkan data historis [21]. Algoritma ini bekerja Dengan memanfaatkan konsep probabilitas untuk mengklasifikasikan data ke dalam kategori tertentu. Dalam konteks minat konsumen pada ponsel, Naive Bayes dapat digunakan untuk memprediksi jenis ponsel yang kemungkinan besar diminati oleh seorang konsumen berdasarkan data historis preferensi mereka atau kelompok serupa.

Misalnya, sebuah ponsel memiliki data historis tentang ponsel yang pernah dibeli atau dicari oleh konsumen. Data ini mencakup informasi seperti merek, harga, kapasitas baterai, kualitas kamera, dan ulasan pelanggan. Naive Bayes akan menggunakan data ini untuk menghitung probabilitas setiap kategori (misalnya, "ponsel premium" atau "ponsel budget-friendly") berdasarkan fitur tertentu.

Cara kerjanya sederhana namun efektif. Algoritma akan menghitung peluang bahwa seorang konsumen memilih kategori ponsel tertentu, berdasarkan pola yang ditemukan di data historis. Sebagai contoh, jika data menunjukkan bahwa konsumen dengan anggaran di bawah Rp3 juta sering memilih ponsel dengan kapasitas baterai besar, Naive Bayes akan memberikan probabilitas tinggi pada kategori ponsel dengan fitur tersebut ketika konsumen baru dengan profil serupa muncul.

Keunggulan utama Naive Bayes adalah kemampuannya untuk bekerja baik meskipun data berukuran besar dan kompleks [22]. Dengan asumsi bahwa setiap

fitur (seperti harga dan merek) dianggap independen satu sama lain, algoritma ini tetap mampu memberikan hasil yang akurat dalam banyak kasus. Hasil prediksi ini dapat digunakan oleh perusahaan untuk memberikan rekomendasi personal kepada konsumen. Misalnya, ketika seorang konsumen mencari ponsel, sistem dapat langsung menampilkan produk yang paling sesuai dengan preferensi mereka berdasarkan analisis data historis menggunakan Naive Bayes [23]. Hal ini membantu meningkatkan pengalaman belanja sekaligus mendorong penjualan.

Mitigasi terhadap kekurangan algoritma, seperti dalam pemilihan fitur, memerlukan pemahaman yang mendalam tentang konteks masalah, data, dan tujuan yang ingin dicapai. Berikut adalah langkah-langkah yang dapat dilakukan. Pertama, mulailah dengan memahami konteks dan domain masalah. Ini penting agar algoritma yang digunakan tidak hanya berdasarkan asumsi statistik, tetapi juga relevan dengan masalah yang dihadapi. Misalnya, dalam analisis kesehatan, fitur seperti usia, berat badan, dan riwayat medis mungkin lebih relevan daripada fitur acak seperti preferensi musik. Kemudian, lakukan eksplorasi data secara menyeluruh untuk mengidentifikasi pola, tren, atau anomali yang ada. Pemahaman ini membantu dalam memilih fitur yang memiliki hubungan langsung dengan variabel target. Setelah itu, gunakan metode berbasis data untuk mengidentifikasi fitur penting. Ini bisa dilakukan dengan teknik seperti analisis korelasi, metode statistik, atau algoritma machine learning seperti pohon keputusan yang mampu menunjukkan pentingnya setiap fitur. Menurut Pradana dan Sugiharti, algoritma ini memiliki kinerja yang kompetitif dibandingkan dengan metode klasifikasi lainnya dan dapat diterapkan dalam berbagai situasi dunia nyata yang kompleks [24].

Nuraeni et al. menunjukkan bahwa Naive Bayes dapat mencapai akurasi 80,33% dalam klasifikasi, yang menunjukkan efektivitasnya dalam analisis data [25]. Lebih lanjut, Rachman dan Handayani menjelaskan bahwa algoritma ini digunakan untuk memprediksi kelancaran pembayaran sewa, menunjukkan kemampuannya dalam pengolahan data historis untuk pengambilan keputusan [26]. Selain itu, Christian et al. menekankan bahwa Naive Bayes dapat digunakan untuk prediksi probabilitas keanggotaan suatu kelas, yang menjadikannya alat yang berguna dalam berbagai aplikasi, termasuk analisis kualitas air [27]. Dengan demikian, Naive Bayes merupakan algoritma yang sederhana namun efektif dalam klasifikasi data.

2.3.1. Prinsip Dasar Metode Naïve Bayes

Prinsip dasar algoritma Naïve Bayes berlandaskan pada Teorema Bayes, yang menghitung probabilitas suatu kelas berdasarkan fitur yang ada. Algoritma ini mengasumsikan bahwa setiap fitur bersifat independen satu sama lain, yang merupakan alasan mengapa disebut "naïve" atau naif Hendri [28]. Dalam penerapannya, Naïve Bayes menghitung probabilitas posterior untuk setiap kelas dan memilih kelas dengan probabilitas tertinggi sebagai hasil klasifikasi [9]. Naïve Bayes dikenal karena kesederhanaannya dan kecepatan dalam proses perhitungan, menjadikannya salah satu algoritma yang populer dalam data mining dan machine learning [29]. Meskipun memiliki kelebihan, algoritma ini juga memiliki kelemahan, terutama dalam hal akurasi ketika fitur tidak independen atau ketika ada hubungan antar fitur [30]. Namun, dengan optimasi yang tepat, seperti pemilihan fitur yang baik, akurasi Naïve Bayes dapat ditingkatkan secara signifikan [31].

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

2.3.2. Kelebihan dan Kekurangan Metode Naive Bayes

1. Kelebihan Metode Naive Bayes

Metode Naive Bayes memiliki sejumlah kelebihan yang menjadikannya populer dalam berbagai aplikasi klasifikasi data [32]. Salah satu keunggulan utamanya adalah kemampuannya untuk bekerja secara efisien dengan jumlah data yang besar, bahkan ketika data memiliki banyak variabel atau atribut [33]. Selain itu, algoritma ini tergolong sederhana dan cepat dalam proses pelatihan maupun prediksi, sehingga sangat cocok digunakan dalam situasi yang membutuhkan analisis waktu nyata. Naive Bayes juga memiliki performa yang cukup baik meskipun asumsi independensi antar fitur sering kali tidak sepenuhnya terpenuhi, serta mampu menangani data kategori maupun numerik dengan baik. Keunggulan lainnya adalah kemampuannya dalam memberikan probabilitas hasil klasifikasi, yang dapat membantu dalam pengambilan keputusan yang lebih terukur dan akurat.

Selain itu, metode Naive Bayes juga dikenal sangat tangguh dalam menangani masalah data yang tidak seimbang, di mana jumlah data dalam satu kelas jauh lebih banyak dibanding kelas lainnya. Hal ini membuatnya cocok digunakan dalam berbagai skenario nyata, seperti deteksi spam, diagnosis medis, hingga analisis perilaku pelanggan. Kemudahan dalam implementasi dan interpretasi hasil menjadikan algoritma ini sebagai pilihan yang ideal untuk peneliti maupun praktisi yang menginginkan solusi klasifikasi yang cepat dan efektif.

Dengan seluruh kelebihannya, Naive Bayes menjadi metode yang layak dipertimbangkan dalam berbagai proyek data mining dan machine learning, khususnya ketika sumber daya komputasi terbatas namun tetap menginginkan hasil yang andal.

2. Kekurangan Metode Naive Bayes

Meskipun memiliki banyak kelebihan, metode Naive Bayes juga memiliki beberapa kekurangan yang perlu diperhatikan. Salah satu kelemahan utamanya adalah asumsi independensi antar fitur, yang jarang benar-benar terjadi dalam data dunia nyata. Dalam banyak kasus, atribut-atribut dalam sebuah dataset saling bergantung satu sama lain, dan asumsi bahwa setiap fitur tidak saling memengaruhi dapat menyebabkan penurunan akurasi model. Hal ini menjadi masalah terutama jika terdapat korelasi kuat antar fitur, karena model tidak dapat menangkap hubungan kompleks tersebut.

Selain itu, Naive Bayes juga kurang efektif ketika berhadapan dengan data numerik yang tidak memiliki distribusi normal, atau jika distribusi kelas tidak seimbang secara ekstrem. Metode ini juga cenderung memberikan performa yang kurang optimal dalam dataset kecil dengan noise tinggi, karena sangat bergantung pada distribusi probabilitas dari setiap kelas. Meskipun hasilnya cepat, model ini tidak dapat melakukan pembelajaran yang mendalam atau membentuk hubungan nonlinear kompleks sebagaimana algoritma seperti Decision Tree atau Random Forest. Oleh karena itu, penting untuk mempertimbangkan konteks dan karakteristik data sebelum memutuskan untuk menggunakan Naive Bayes dalam sebuah penelitian atau proyek klasifikasi.

2.3.3. Evaluasi Metode Naive Bayes

Evaluasi model Naïve Bayes melibatkan beberapa aspek penting, termasuk akurasi, presisi, recall, dan F1-score. Ketika menggunakan algoritma untuk memprediksi minat konsumen pada ponsel, kita perlu mengevaluasi kinerjanya dengan mengukur akurasi, presisi, dan recall. Ketiga metrik ini membantu kita memahami sejauh mana algoritma mampu membuat prediksi yang benar dan relevan dengan kebutuhan bisnis.

Akurasi adalah ukuran seberapa sering algoritma membuat prediksi yang benar dibandingkan dengan semua prediksi yang dibuat. Misalnya, jika algoritma memprediksi apakah seseorang akan tertarik pada ponsel tertentu, akurasi menunjukkan persentase prediksi benar dari keseluruhan data. Namun, akurasi bisa menyesatkan jika data tidak seimbang, seperti ketika sebagian besar orang dalam dataset memang tidak tertarik pada produk.

Presisi berfokus pada kualitas prediksi positif. Dalam konteks ini, presisi mengukur seberapa sering algoritma benar-benar tepat ketika mengatakan seorang konsumen akan tertarik pada ponsel. Misalnya, jika algoritma memprediksi 100 orang akan tertarik, dan hanya 70 dari mereka benar-benar tertarik, maka presisi adalah 70%. Ini penting jika bisnis ingin meminimalkan kesalahan dalam menargetkan konsumen yang tidak relevan.

Recall mengukur kemampuan algoritma untuk menangkap semua konsumen yang benar-benar tertarik pada ponsel, dibandingkan dengan jumlah total orang yang seharusnya teridentifikasi. Jika ada 200 orang dalam dataset yang benar-benar tertarik pada ponsel, dan algoritma hanya berhasil memprediksi 150 dari mereka,

recall-nya adalah 75%. Recall sangat penting ketika perusahaan ingin memastikan tidak melewatkan konsumen potensial.

Ketiga metrik ini sering digunakan bersama untuk mendapatkan gambaran yang lebih lengkap tentang kinerja algoritma. Jika presisi dan recall tidak seimbang, kita bisa menggunakan metrik lain seperti F1-score untuk mengombinasikan keduanya. Dengan memahami metrik-metrik ini, tim pemasaran atau data dapat mengidentifikasi kekuatan dan kelemahan algoritma, serta membuat penyesuaian yang diperlukan untuk meningkatkan prediksi minat konsumen secara lebih akurat.

Kita bisa menggunakan contoh nyata untuk memahami bagaimana evaluasi model Naive Bayes diterapkan menggunakan confusion matrix dan metrik evaluasi seperti akurasi, presisi, dan recall. Misalnya, kita memiliki dataset yang berisi informasi tentang harga dan fitur ponsel, serta label apakah konsumen tertarik atau tidak. Model Naive Bayes digunakan untuk memprediksi minat konsumen berdasarkan data ini.

2.4. Model Klasifikasi

Model klasifikasi merupakan salah satu teknik dalam machine learning yang digunakan untuk mengelompokkan data ke dalam kategori atau kelas tertentu berdasarkan atribut-atribut yang dimilikinya. Tujuan utama dari model klasifikasi adalah untuk memprediksi label atau kelas dari data baru dengan menggunakan pola yang telah dipelajari dari data sebelumnya (data training). Model ini bekerja dengan mencari hubungan antara fitur-fitur (atribut) input dengan kelas target, sehingga mampu mengenali karakteristik masing-masing kategori.

Berbagai algoritma dapat digunakan dalam membangun model klasifikasi, seperti Decision Tree, Naive Bayes, K-Nearest Neighbor (KNN), dan Support Vector Machine (SVM). Masing-masing algoritma memiliki cara kerja yang berbeda dalam membentuk model dan memiliki kelebihan serta kekurangan tersendiri. Misalnya, algoritma Decision Tree bekerja dengan membuat struktur pohon berdasarkan pembagian atribut yang paling optimal, sementara Naive Bayes menggunakan prinsip probabilitas dan asumsi independensi antar atribut. Pemilihan algoritma yang tepat sangat bergantung pada jenis data, tujuan analisis, dan konteks penggunaannya.

Proses pembangunan model klasifikasi meliputi beberapa tahapan, dimulai dari pengumpulan data, pembersihan dan seleksi data (preprocessing), pelatihan model menggunakan data training, dan pengujian model menggunakan data testing. Evaluasi performa model dilakukan menggunakan metrik seperti akurasi, presisi, recall, dan F1-score. Evaluasi ini bertujuan untuk mengetahui seberapa baik model dalam melakukan klasifikasi terhadap data baru. Semakin tinggi nilai evaluasi tersebut, maka semakin baik pula performa model klasifikasi yang dibangun.

Model klasifikasi memiliki peran penting dalam berbagai bidang, seperti kesehatan (untuk memprediksi penyakit), keuangan (untuk deteksi penipuan), pemasaran (untuk segmentasi pelanggan), dan lainnya. Dalam konteks penelitian ini, model klasifikasi digunakan untuk mengidentifikasi preferensi masyarakat terhadap pemilihan merek HP berdasarkan sejumlah atribut tertentu. Dengan pendekatan ini, hasil analisis tidak hanya bermanfaat dalam memahami perilaku

konsumen, tetapi juga membantu pelaku usaha dan pembuat kebijakan dalam menyusun strategi yang lebih tepat sasaran.

2.5. Alat Bantu Program Aplikasi RapidMiner

RapidMiner adalah salah satu perangkat lunak berbasis data science yang dirancang untuk memudahkan proses analisis data, termasuk dalam membangun model klasifikasi, prediksi, dan eksplorasi data. Alat bantu ini bersifat visual dan drag-and-drop, sehingga sangat cocok digunakan oleh peneliti atau praktisi yang tidak memiliki latar belakang pemrograman yang kuat. Dengan tampilan antarmuka yang intuitif, pengguna dapat menyusun proses analisis hanya dengan menyambungkan blok-blok proses (operator) secara logis.

RapidMiner mendukung berbagai teknik analisis, mulai dari preprocessing data, transformasi, pemodelan, hingga evaluasi hasil. Salah satu keunggulan RapidMiner adalah tersedianya banyak algoritma klasifikasi seperti Naive Bayes, Decision Tree, K-NN, dan lain-lain yang dapat diakses langsung tanpa perlu menulis kode. Selain itu, proses integrasi data dari berbagai sumber (file Excel, database, atau CSV) dapat dilakukan dengan mudah, sehingga mempersingkat waktu pengolahan data.

Dalam proses klasifikasi, RapidMiner menyediakan alur kerja (workflow) yang jelas, mulai dari pemilihan atribut, pembagian data menjadi training dan testing, pelatihan model, hingga visualisasi hasil evaluasi. Output dari setiap langkah dapat dimonitor secara langsung, sehingga memudahkan dalam melakukan perbaikan atau penyesuaian parameter. Fitur ini sangat membantu dalam

mengembangkan model yang akurat dan dapat dipercaya untuk pengambilan keputusan berbasis data.

RapidMiner juga memiliki fitur untuk mengevaluasi performa model secara mendalam, seperti confusion matrix, akurasi, precision, recall, dan F1-score. Hasil evaluasi tersebut memungkinkan pengguna untuk menilai efektivitas model yang dibangun dalam mengklasifikasikan data secara tepat. Dengan kemampuannya yang lengkap dan fleksibel, RapidMiner menjadi salah satu pilihan utama dalam pengolahan data, khususnya dalam penelitian yang membutuhkan analisis klasifikasi berbasis algoritma seperti Naive Bayes.

2.6. Kelebihan Penelitian

Penelitian ini memiliki sejumlah kelebihan yang membuatnya relevan dan bermanfaat, baik dari segi akademis maupun praktis. Salah satu kelebihan utama adalah penerapan algoritma Naive Bayes dalam konteks pemilihan merek HP di masyarakat Labuhanbatu, yang jarang dieksplorasi dalam studi-studi sebelumnya. Dengan mengombinasikan pendekatan data mining dan analisis perilaku konsumen, penelitian ini berhasil memberikan kontribusi ilmiah dalam memahami preferensi masyarakat melalui pendekatan berbasis data.

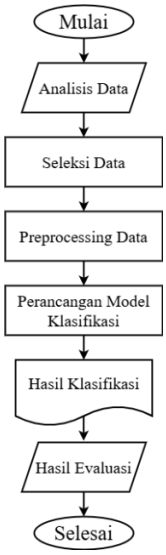
Selain itu, penelitian ini memanfaatkan alat bantu seperti RapidMiner, yang memungkinkan proses analisis dilakukan secara sistematis, efisien, dan visual. Penggunaan perangkat ini mempermudah proses klasifikasi dan evaluasi, serta menghasilkan model prediksi yang lebih akurat dan dapat diuji secara langsung. Hal ini menjadikan penelitian tidak hanya relevan secara konseptual, tetapi juga

kuat dari segi metodologi, karena didukung oleh teknologi yang telah terbukti dalam dunia data science.

Kelebihan lain dari penelitian ini adalah pengambilan data yang berfokus pada masyarakat lokal, yaitu di wilayah Labuhanbatu. Hal ini memberikan nilai tambah karena hasil penelitian mencerminkan kondisi nyata di lapangan dan dapat menjadi dasar dalam pengambilan keputusan strategis oleh pelaku usaha atau pihak terkait lainnya. Penelitian ini juga membuka peluang bagi pengembangan strategi pemasaran yang lebih efektif dan sesuai dengan kebutuhan masyarakat setempat.

Terakhir, penelitian ini memberikan manfaat langsung kepada konsumen dengan menyediakan informasi tentang tren dan kecenderungan preferensi terhadap merek HP. Hasil klasifikasi dapat dijadikan panduan dalam memilih produk yang sesuai dengan kebutuhan dan anggaran. Dengan demikian, penelitian ini tidak hanya memberikan wawasan akademik, tetapi juga solusi praktis yang dapat diterapkan dalam kehidupan sehari-hari masyarakat.

2.7. Kerangka Kerja Penelitian



Gambar 2. 2. Kerangka Kerja Penelitian

Berikut adalah penjelasan untuk setiap tahapan dalam flowchart kerangka kerja penelitian yang ditampilkan:

1. Analisis Data, Pada tahap ini dilakukan pengamatan terhadap data mentah untuk memahami karakteristik serta struktur data yang tersedia. Tujuannya adalah untuk mengidentifikasi informasi penting yang relevan untuk proses klasifikasi.
2. Seleksi Data, Tahapan ini bertujuan untuk memilih atribut-atribut yang dianggap paling berpengaruh terhadap target klasifikasi. Dengan seleksi data yang tepat, proses analisis menjadi lebih fokus dan efisien.
3. Preprocessing Data, Data yang telah dipilih kemudian dibersihkan dan disiapkan agar layak untuk dianalisis lebih lanjut, termasuk penanganan data kosong atau duplikat. Proses ini penting agar model klasifikasi dapat bekerja secara optimal.
4. Perancangan Model Klasifikasi, Pada tahap ini dibangun model klasifikasi menggunakan algoritma tertentu, dalam hal ini Naive Bayes. Model dirancang berdasarkan data yang telah diproses untuk mempelajari pola yang ada.
5. Hasil Klasifikasi, Model yang telah dibangun digunakan untuk melakukan klasifikasi pada data baru atau data uji. Hasil klasifikasi ini menunjukkan prediksi model terhadap kategori data berdasarkan pola yang dipelajari.
6. Hasil Evaluasi, Evaluasi dilakukan untuk menilai kinerja model klasifikasi, biasanya menggunakan metrik seperti akurasi, presisi, dan recall. Tahapan ini penting untuk mengetahui seberapa baik model dalam melakukan prediksi dan klasifikasi.