BAB II
LANDASAN TEORI

2.1. Kepuasan Pelanggan

Kepuasan pelanggan adalah bagian yang berhubungan dengan penciptaan nilai pelanggan. Karena terciptanya kepuasan pelanggan berarti memberikan manfaat bagi perusahaan yaitu, diantaranya hubungan antara perusahaan dengan pelanggannya menjadi harmonis, memberikan dasar yang baik atau terciptanya kepuasan pelanggan serta membentuk suatu rekomendasi dari mulut ke mulut yang menguntungkan bagi perusahaan, sehingga timbul Puas dari pelanggan untuk membeli atau menggunakan jasa perusahaan tersebut.

Kepuasan adalah perasaan senang atau kecewa seseorang yang muncul setelah membandingkan kinerja (hasil) produk yang difikirkan terhadap kinerja (hasil) yang diharapkan.Jika kinerja di bawah harapan, maka pelanggan tidak puas.Jika kinerja memenuhi harapan, pelanggan puas.Jika kinerja melebihi harapan, pelanggan amat puas atau senang[1]. kepuasan adalah perbedaan antara harapan dan unjuk kerja.Kepuasan pelanggan selalu didasarkan pada upaya peniadaan atau menyempitkan gap antar harapan dan kinerja.

2.2. IndiHome

Perkembangan teknologi ruang online saat ini semakin mempermudah pemenuhan kebutuhan informasi masyarakat. Salah satu teknologi jaringan yang paling cepat berkembang adalah teknologi jaringan media transport nirkabel atau wireless. Komunikasi nirkabel adalah gelombang radio yang dapat ditransmisikan

ke area mana pun yang dapat dijangkau oleh gelombang radio ini. Media transmisi data nirkabel yang populer adalah Wireless Fidelity (WiFi)[2].

Salah satu produk layanan PT Telekomunikasi Indonesia,Tbk Branch Consumer Rantauprapat adalah Digital Home, atau IndiHome. Produk ini mencakup layanan komunikasi dan data seperti telepon rumah (suara), internet (Internet on Fiber atau High Speed Internet), dan televisi interaktif (Usee TV Cable dan IP TV). Telkom menggambarkan IndiHome sebagai tiga layanan dalam satu paket (3-in-1) karena pelanggan dapat menikmati tayangan TV berbarengan dengan internet. Indihome resmi dirilis pada awal tahun 2015. Salah satu inisiatif utama Telkom adalah IndiHome. Telkom menggandeng berbagai pengembangan teknologi komunikasi untuk membangun rumah digital selama penyelenggaraannya. Pelayanan Indihome hanya dapat digunakan di rumah di mana jaringan serat optik Telkom (FTTH) tersedia dan di mana kabel tembaga masih digunakan.

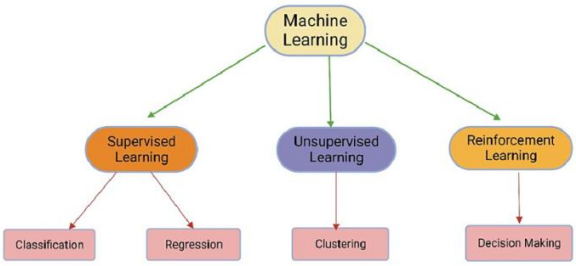
2.3. Machine Learning

Machine learning (Pembelajaran Mesin) adalah suatu mesin yang dikembangkan dengan kemampuan komputer untuk melakukan pembelajaran sendiri tanpa ada arahan. Pembelajaran mesin (ML) adalah studi ilmiah mengenai algoritma dan model statistik sistem komputer yang digunakan untuk melakukan tugas tertentu tanpa diprogram secara eksplisit. Pada praktiknya, algoritma Machine Learning bertugas menganalisis pola dalam data untuk menghasilkan prediksi atau klasifikasi berdasarkan pola tersebut. Sebagai teknologi yang terus berkembang, ML telah menunjukkan potensi besar dalam memproses data,

meningkatkan efisiensi operasional, serta mendukung inovasi di berbagai bidang, termasuk analisis kepuasan pelanggan [3]. Dalam analisis kepuasan pelanggan, Machine Learning memainkan peran penting untuk menganalisis ulasan pelanggan dan komentar media sosial secara otomatis. Algoritma seperti Naïve Bayes dan K-Nearest Neighbor digunakan untuk mengklasifikasikan sentimen pelanggan ke dalam kategori positif, negatif, atau netral. Proses ini membantu perusahaan memahami kebutuhan pelanggan dan mengambil langkah strategis berdasarkan data yang dihasilkan secara real-time [4]; [5]. Machine learning sebagai bagian dari data mining yang berfokus pada pembuatan model berbasis algoritma yang dapat mengidentifikasi pola dan melakukan prediksi atau klasifikasi. Machine Learning dapat diklasifikasikan menjadi tiga jenis utama, yaitu:

1. Pembelajaran Terawasi (Supervised Learning): Pada jenis ini, algoritma dilatih menggunakan dataset berlabel, yang berarti setiap data input memiliki pasangan output yang diketahui. Tujuannya adalah untuk memprediksi hasil pada data baru berdasarkan pola yang ditemukan dalam data pelatihan [6]. Penelitian ini menggunakan supervised learning untuk klasifikasi tingkat kepuasan pelanggan IndiHome dengan algoritma Naïve Bayes dan K-Nearest Neighbor.
2. Pembelajaran Tak Terawasi (Unsupervised Learning): Algoritma ini bertujuan menemukan pola tersembunyi dalam data yang tidak memiliki label. Contoh aplikasi meliputi pengelompokan pelanggan dan analisis segmentasi pasar [7].

3. Pembelajaran Penguatan (Reinforcement Learning): Model belajar dengan berinteraksi langsung dengan lingkungan untuk memaksimalkan reward tertentu. Metode ini sering digunakan dalam pengendalian sistem dan pengembangan robot.

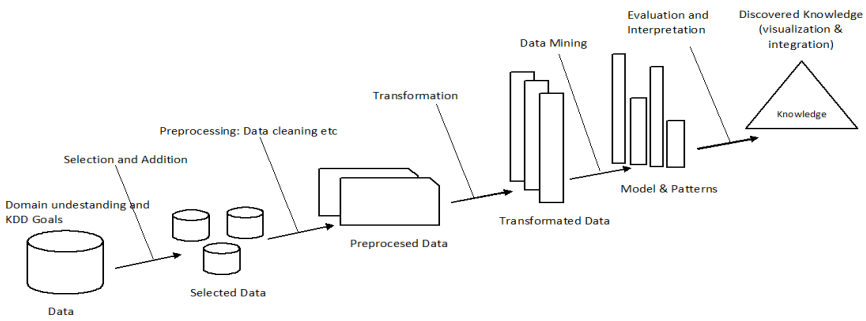


Sumber: [8]

Gambar 2.1 Diagram Alur Jenis-jenis Machine Learning

2.4. Knowledge Discovery in Database

Knowledge Discovery in Databases (KDD) adalah proses sistematis untuk menemukan pengetahuan atau pola yang berguna dari kumpulan data besar. Proses ini mencakup beberapa tahap utama, yaitu: pembersihan data (data cleaning), integrasi data, seleksi data, transformasi data, penambangan data (data mining), evaluasi pola, dan presentasi hasil. KDD banyak digunakan dalam berbagai bidang seperti bisnis, kesehatan, dan pendidikan untuk membantu pengambilan keputusan yang lebih akurat dan efisien.



Gambar 2.2 Diagram Alur Jenis-jenis Machine Learning

Teknik data mining disebut dengan nama Knowledge Discovery Database (KDD). Database Discovery Knowledge (KDD) adalah istilah penerapan metode ilmiah pada data mining. Dalam ini, data mining merupakan langkah dalam proses Knowledge Discovery Database (KDD). Berikut merupakan penjelasaan tahapan dari proses Knowlegde Discovery in Database (KDD):

1. Informasi Seleksi (selection), Pada tahap ini, data yang relevan untuk analisis diidentifikasi dan dipilih. Pemilihan data yang tepat sangatlah penting karena kualitas dan relevansi data mempengaruhi hasil akhir dari proses Knowlegde Discovery in Database (KDD).
2. Persiapan Data (Pembersihan/Preprocessing), Pemrosesan dan pembersihan data adalah dua tahapan krusial dalam siklus Knowledge Discovery In Database (KDD). Proses ini bertujuan untuk memastikan bahwa data yang digunakan dalam analisis atau pemodelan tidak hanya lengkap dan konsisten, tetapi juga sesuai dan siap digunakan dalam model analisis. Tanpa pemrosesan dan pembersihan data yang tepat, hasil analisis atau model yang dibangun mungkin menyesatkan atau tidak valid.
3. Transformasi Data, Setelah data dibersihkan, langkah selanjutnya adalah mentransformasikan data. Transformasi data bertujuan untuk mengubah data yang telah diseleksi dan dibersihkan kedalam format atau bentuk yang lebih sesuai untuk dianalisis lebih lanjut. Pada tahap ini, menggunakan teknik tertentu, seperti data mining.

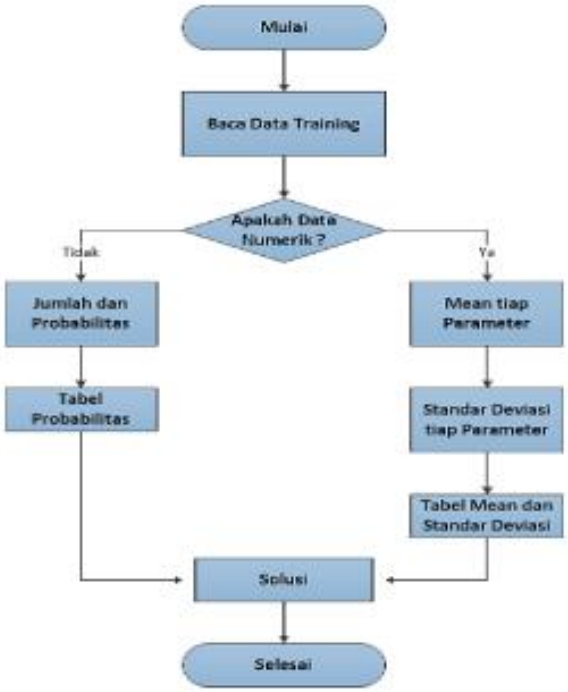
4. Data Mining, Data mining adalah langkah utama KDD. Pada fase ini diterapkan teknik analisis algoritma untuk menemukan pola atau informasi tersembunyi dalam data. Data mining melibatkan penggunaan berbagai teknik statistik dan pembelajaran mesin untuk menganalisis data dan menemukan hubungan atau pola yang sebelumnya tidak diketahui.
5. Interpretasi/Evaluasi, Setelah pola atau model sudah ditemukan maka langkah selanjutnya adalah evaluasi. Pada fase ini model yang ditemukan diuji untuk mengevaluasi efektivitasnya dalam mendeskripsikan data hasil yang dinginkan. Evaluasi bertujuan untuk mengetahui apakah pola atau model yang ditemukan cukup valid dan layak digunakan dalam pengambilan Keputusan.

2.5. Algoritma Naive bayes

2.5.1. Pengertian Naïve Bayes

Metode Naive Bayes merupakan salah satu teknik dalam data mining karena memiliki kemampuan untuk melakukan analisis dan pemrosesan data dalam jumlah besar, serta menghasilkan model yang dapat digunakan untuk melakukan prediksi atau klasifikasi. Naïve Bayes adalah algoritma klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes untuk menghitung probabilitas suatu kelas berdasarkan atribut-atribut yang dimiliki oleh data. Algoritma ini disebut "naïve" karena mengasumsikan independensi antar-atribut dalam dataset, yang artinya setiap atribut dianggap tidak saling memengaruhi.Kesederhanaan algoritma Naïve Bayes menjadikannya populer dalam analisis data karena efisiensinya dalam menangani dataset besar serta kemampuan menghasilkan

prediksi yang akurat dalam waktu yang relatif singkat [9]. Selain itu, algoritma ini cocok digunakan untuk atribut kategorikal, sehingga sering diterapkan pada berbagai kasus klasifikasi, Seperti analisis sentimen, klasifikasi teks, dan deteksi spam [10].



Sumber:[11]

Gambar 2.3 Diagram Alur *Naïve Bayes*

2.5.2. Formula Matematika *Naïve Bayes*

Teorema Bayes menggunakan probabilitas untuk memprediksi kelas berdasarkan rumus:

$$\mathbf{P\ (A\mid B):}\ \frac{\boldsymbol{P\ (B\mid A)\ P\ (A)}}{\boldsymbol{P\ (B)}}$$

Keterangan :

A : Hipotesis data A (kelas tertentu)

B : Data dengan kelas yang tidak diketahui

P (A | B) : Probabilitas hipotesis berdasarkan kondisi B

P (A) : Kemungkinan hipotesis A

P (B | A) : Probabilitas B ketika kondisi A

P (B) : Probabilitas dari B

2.5.3. Kelebihan dan Kekurangan *Naïve Bayes*

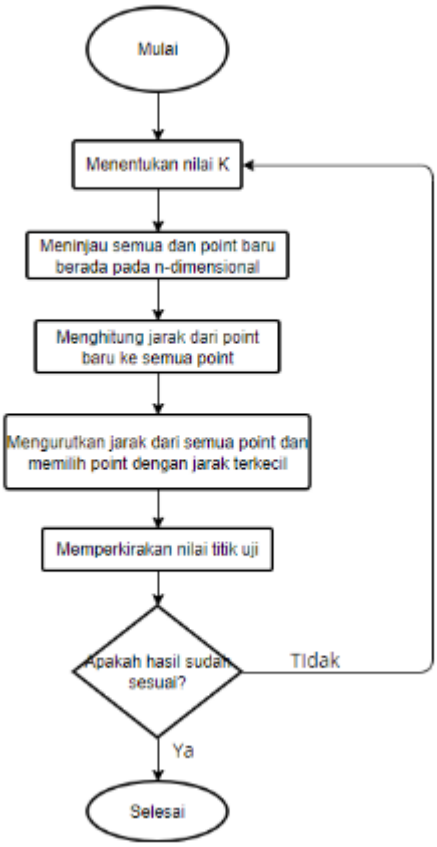
Algoritma ini memiliki kelebihan dalam hal kecepatan komputasi dan efisiensi pada dataset besar, terutama pada klasifikasi berbasis teks [12]. Namun, asumsi independensi antar-fitur menjadi kelemahannya ketika data memiliki hubungan antar-variabel[13].

2.6. Algoritma K-Nearest Neighbor (KNN)

2.6.1. Pengertian KNN

Metode K-Nearest Neighbors (K-NN) merupakan salah satu teknik dalam data mining yang digunakan untuk menganalisis dan mengklasifikasikan data dengan menentukan kelas suatu data baru berdasarkan kedekatan dengan tetangga-tetangga terdekatnya dalam ruang multidimensi. Dalam algoritma ini, setiap data baru diklasifikasikan dengan cara mencari k tetangga terdekatnya dari dataset yang sudah ada, kemudian memutuskan kategori data berdasarkan mayoritas kelas dari tetangga tersebut[14]. KNN terkenal karena kesederhanaannya, di mana algoritma ini tidak memerlukan model pelatihan secara eksplisit. Sebagai gantinya, KNN menggunakan data pelatihan secara langsung untuk melakukan klasifikasi berdasarkan jarak antar data. Metode ini sangat fleksibel dan mampu menangani dataset, termasuk data yang tidak

memiliki distribusi linier [15]. Keunggulan lainnya adalah kemampuan KNN untuk bekerja dengan berbagai jenis data, seperti data numerik maupun kategorikal, dan dapat diterapkan pada berbagai aplikasi, seperti klasifikasi kelayakan bangunan sekolah[16], deteksi dini penyakit stroke[17], dan



penggolongan jurusan siswa [18].

Sumber: [19]

Gambar 2.4 Gambar Diagram Alur Proses *K-Nearest Neighbor*

Visualisasi ini menggambarkan langkah-langkah kerja algoritma K-Nearest Neighbors (KNN) dalam klasifikasi data, dimulai dari pengukuran jarak hingga proses klasifikasi.

1. Pengukuran Jarak: Langkah pertama adalah mengukur jarak antara data baru yang ingin diklasifikasikan dengan data latih. Jarak ini biasanya dihitung menggunakan rumus Euclidean distance, tetapi bisa juga menggunakan metode lain seperti Manhattan atau Minkowski, tergantung pada kebutuhan.
2. Menentukan K Tetangga Terdekat: Setelah jarak dihitung, algoritma memilih K tetangga terdekat, yaitu titik data dari dataset latih yang memiliki jarak paling kecil terhadap data baru.
3. Klasifikasi Berdasarkan Mayoritas: Setelah mendapatkan K tetangga terdekat, algoritma mengidentifikasi kelas yang paling sering muncul di antara K tetangga tersebut. Kelas mayoritas ini kemudian menjadi label prediksi untuk data baru.
4. Output Klasifikasi: Hasilnya adalah klasifikasi yang menunjukkan kelas atau kategori yang diprediksi berdasarkan mayoritas label dari K tetangga terdekat.

2.6.2. Formula Matematika KNN

KNN menggunakan jarak Euclidean untuk mengukur kedekatan antara data:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_{training}^{i} - y_{testing}^{i})}$$

Keterangan:

d (x, y) : Jarak

$x^i_{training}$: Data Training

$y^i_{testing}$: Data Testing

i : Variabel Data

n : Dimensi Data

2.6.3. Kelebihan dan Kekurangan KNN

KNN sangat fleksibel dalam menangani pola data, terutama pada dataset yang tidak memiliki distribusi linier. Namun, kelemahannya terletak pada kebutuhan komputasi tinggi, terutama jika nilai k tidak dioptimalkan, serta sensitivitas terhadap outlier.

2.7. Evaluasi Model

Pada tahap Evaluasi Model, Tahapan ini bertujuan untuk menilai kinerja model dalam mengklasifikasikan sentimen pelanggan. Hasil evaluasi akan membantu peneliti untuk memahami kelebihan dan kekurangan dari kedua algoritma yang digunakan, serta memberikan insight mengenai kemungkinan perbaikan model yang lebih baik di masa depan. Pada penelitian ini Confusion matrix Digunakan untuk menganalisis distribusi prediksi model terhadap data aktual. Confusion matrix memberikan wawasan tentang kesalahan yang terjadi dalam klasifikasi, seperti false positives dan false negatives. True Positive (TP)

kasus di mana model memprediksi "positif" dan memang benar positif.False Positive (FP) kasus di mana model memprediksi "positif", tetapi sebenarnya negative, False Negative (FN): Kasus di mana model memprediksi "negatif", tetapi sebenarnya positif , True Negative (TN) kasus di mana model memprediksi "negatif" dan memang benar negatif. Metrik yang digunakan untuk evaluasi meliputi:

Akurasi : Mengukur sejauh mana model dapat memprediksi dengan benar.

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

Presisi : Mengukur seberapa tepat model dalam mengidentifikasi sentimen

$$Presisi = \frac{TP}{TP + FP}$$

Recall : Mengukur kemampuan model dalam menemukan semua kasus sentimen positif atau negatif

$$Recall = \frac{TP}{TP + FN}$$

F1-Score : Merupakan rata-rata harmonis dari presisi dan recall, memberikan Gambaran umum dari kinerja model.

$$F1 - Score = 2 \times \frac{Presisi \times Recall}{Presisi + Recall}$$

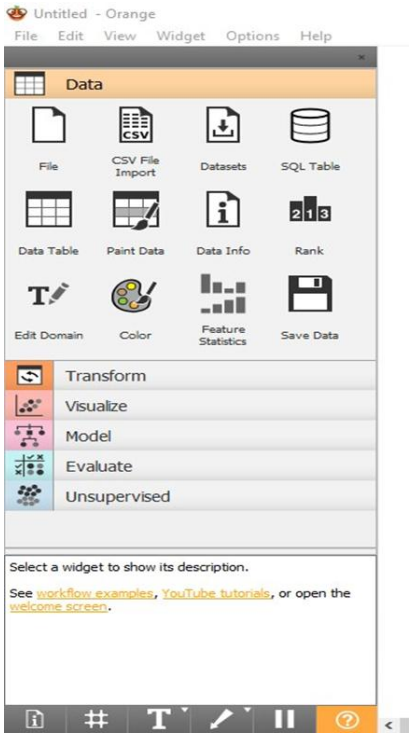
2.8. Model Klasifikasi

Model klasifikasi adalah salah satu teknik dalam data mining yang digunakan untuk mengelompokkan data ke dalam kategori atau kelas tertentu

berdasarkan atribut atau fitur yang ada. Dalam model ini, data yang telah dilabeli digunakan untuk melatih algoritma agar dapat memprediksi kelas dari data yang belum dilabeli[20]. Beberapa metode yang umum digunakan dalam klasifikasi adalah K-Nearest Neighbors (KNN), Naive Bayes, dan Decision Tree, yang masing-masing memiliki keunggulan dan aplikasi yang berbeda[21]. Proses klasifikasi ini biasanya dimulai dengan pengumpulan data, pra-pemrosesan untuk membersihkan dan menyiapkan data, serta pemilihan fitur yang relevan[22]. Setelah model dilatih, dilakukan evaluasi untuk mengukur akurasi dan keandalan prediksi yang dihasilkan.

2.9. Alat Bantu Program Aplikasi Orange

Aplikasi Orange adalah teknologi pembelajaran mesin sumber terbuka atau perangkat lunak penambangan data. Orange dapat digunakan untuk menganalisis dan memvisualisasikan data eksplorasi. Ini menyediakan platform untuk pemilihan eksperimental, pemodelan prediktif, dan sistem rekomendasi dan dapat digunakan untuk penelitian genomik, biomedis, bioinformatika, dan pendidikan. Orange selalu diistimewakan dalam hal inovasi, kualitas, atau keandalan. Orange memudahkan pengguna untuk bermain dengan data sumber terbuka dan melakukan alur kerja analisis data secara intuitif. Orange Data menyediakan beberapa utilitas untuk mencari data informasi kata dominan dari konten status dan komentar/ulasan yang akan menghasilkan cloud view secara perlahan utilitas Orange.



Gambar 2.5 *Interface Orange*

Teknik penambangan data membantu mengungkap pengetahuan tersembunyi dalam kelompok data yang dapat digunakan untuk menganalisis dan memprediksi data di masa depan. Klasifikasi adalah metode mengekstraksi rekaman berlabel kelas untuk sekumpulan instance yang tidak terklasifikasi[23]. Sejarah orange versi 3.0 pada tahun 2015 menyertakan komponen inti dalam C++ dengan pembungkus dalam python. Dari versi ini dan seterusnya orange menggunakan pustaka sumber terbuka python umum untuk komputasi ilmiah seperti mumpy scipy dan scikit-leam, sementara pengguna grafis beroperasi dalam kerangka Qt lintas platform. Dan pada tahun 2016 orange versi 3.3 pengembangan menggunakan siklus rilis stabil. Implementsi orange yaitu: orange dapat di

implementasikan untuk memprediksi saham, untuk klasifikasi kelulusan mahasiswa dengan model k-nearest neighbor. implementasi data mining menentukan pola hidup sehat bagi pengguna kb menggunakan algoritma adaboost orange[24]. Kelebihan dan kekurangan Orange. kelebihannya yaitu penggunaan sistem ini mudah dipahami dan tidak perlu menggunakan codingan sedangkan kekurangannya yaitu tetap memerlukan codingan untuk hal tertentu.

2.10. Metode Penelitian

2.10.1. Penelitian Terdahulu

Dibawah ini adalah referensi yang penulis gunakan untuk mendukung dan sebagai landasan pembuatan skripsi. Yaitu sebagai berikut:

Table 2. 1. Tabel Penelitian Terdahulu

Referensi penelitian	1
Judul	Analisis Sentimen Kepuasan Customer Menggunakan Naïve Bayes
Nama penulis	Nurhaliza Bin Aras et al
Tahun	2023
Hasil	Akurasi hingga 89.73% dalam klasifikasi sentimen positif, negatif, dan netral.
Referensi penelitian	2
Judul	Implementasi Machine Learning untuk Analisis Kepuasan Pelanggan Aplikasi KAI Access
Nama penulis	Febrina Tesalonika et al

Tahun	2023
Hasil	SVM memiliki akurasi tertinggi dibandingkan KNN dan Decision Tree
Referensi penelitian	3
Judul	Analisis Sentimen Kepuasan Pemangku Kepentingan Menggunakan Naïve Bayes dan KNN
Nama penulis	Ni Luh Ratniasih et al
Tahun	2023
Hasil	Naïve Bayes lebih akurat (91.13%) dibandingkan KNN (83.06%).

2.11. Kelebihan Penelitian

Penelitian ini memiliki beberapa kelebihan yang membedakannya dari penelitian sebelumnya. Salah satu kelebihan utama adalah fokus pada dataset lokal, khususnya layanan broadband IndiHome, yang memberikan lebih mendalam dan relevansi yang lebih tinggi terhadap preferensi dan perilaku masyarakat pengguna layanan tersebut. Analisis sentimen yang dilakukan pada data lokal dapat memberikan wawasan yang lebih terperinci mengenai kepuasan pelanggan dengan mempertimbangkan faktor-faktor spesifik yang mungkin tidak terdeteksi dalam analisis yang lebih umum. Penelitian sebelumnya yang menggunakan algoritma Naïve Bayes untuk menganalisis ulasan aplikasi lokal, seperti yang dilakukan oleh [25]dan[26], menunjukkan bahwa data lokal lebih

mampu mencerminkan kepuasan pengguna secara lebih akurat karena lebih relevan dengan kondisi sosial dan budaya setempat.

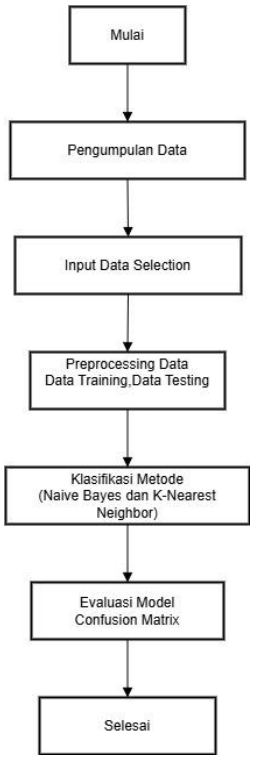
Penelitian ini tidak hanya memanfaatkan dataset lokal, tetapi juga melakukan perbandingan langsung antara algoritma Naïve Bayes dan K-Nearest Neighbor dalam analisis sentimen terhadap ulasan pelanggan IndiHome. Dengan demikian, penelitian ini berfokus pada pengujian efektivitas kedua algoritma tersebut dalam memprediksi tingkat kepuasan pelanggan berdasarkan data lokal yang lebih spesifik. Seperti yang ditunjukkan dalam penelitian oleh [27], perbandingan algoritma dalam analisis sentimen dapat memberikan wawasan yang lebih jelas mengenai kelebihan dan kekurangan masing-masing algoritma yang berbeda.

Kelebihan lain dari penelitian ini adalah kemampuan untuk mengembangkan model yang lebih sesuai dengan karakteristik bahasa dan budaya setempat. Penggunaan dataset lokal memungkinkan peneliti untuk memperhitungkan dinamika sosial dan budaya yang mempengaruhi persepsi pelanggan terhadap layanan broadband IndiHome. Sebagai contoh, analisis sentimen terhadap brand lokal di e-commerce menggunakan pendekatan deep learning yang melibatkan data lokal telah menunjukkan peningkatan akurasi, karena model yang dikembangkan dapat lebih baik menangkap konteks lokal yang khas [28].

Dengan demikian, penelitian ini memberikan kontribusi signifikan dalam mengembangkan model analisis sentimen yang lebih efisien dan akurat dengan menggunakan data yang relevan dan terfokus pada pengguna layanan broadband

lokal, yaitu IndiHome. Selain meningkatkan relevansi hasil analisis, penelitian ini juga berfokus pada pemahaman yang lebih mendalam tentang dinamika sosial yang ada di masyarakat Indonesia, yang pada gilirannya dapat membantu meningkatkan pengalaman pengguna dan kepuasan pelanggan IndiHome.

2.12. Kerangka Kerja Penelitian



Gambar 2. 6. Kerangka Kerja Penelitian

Kerangka kerja penelitian ini terdiri dari tahapan berikut:

1. Pengumpulan data dalam penelitian ini dilakukan dengan observasi secara langsung dan menyebarkan kuesioner kepada pelanggan wifi IndiHome guna mengetahui tingkat kepuasan mereka terhadap layanan IndiHome. Data yang dikumpulkan menggunakan 5 variabel seperti; Kekuatan sinyal, penggunaan, harga berlangganan, jumlah kapasitas user, dan kategori..

2. Selection Data, Tahap ini melibatkan pemilihan data yang relevan untuk penelitian, yaitu data pelanggan wifi IndiHome di Rantauprapat.Data yang dipilih mencakup atribut seperti Kekuatan sinyal, Harga berlangganan, Penggunaan, Jumlah kapasitas user dan Kategori. Data ini akan menjadi dasar dalam proses klasifikasi kepuasan pelanggan wifi IndiHome.
3. Preprocessing Data, Pada tahap ini, data yang telah dipilih akan diproses untuk memastikan kualitas dan kelengkapannya. Proses ini meliputi pembersihan data dari nilai yang hilang atau tidak valid, normalisasi data untuk menyamakan skala, dan transformasi data jika diperlukan agar lebih sesuai untuk analisis. Preprocessing penting untuk meningkatkan akurasi hasil klasifikasi.
4. Klasifikasi Metode Naive Bayes dan K-Nearest Neighbor, Data yang telah diproses akan dianalisis menggunakan metode Naive Bayes dan KNN. Naive Bayes akan mengklasifikasifikasikan data berdasarkan probabilitas teorema bayes dan KNN akan mengklasifikasikan data berdasarkan jarak terdekat dari data baru terhadap data yang sudah ada,. Proses ini bertujuan untuk mengklasifikasikan kepuasaan pelanggan wifi IndiHome
5. Evaluasi Model, Hasil klasifikasi dari kedua metode akan dievaluasi untuk mengukur performa mereka. Evaluasi dilakukan dengan confusion matrix : menghitung metrik seperti akurasi, presisi, dan recall untuk menentukan sejauh mana metode Naive Bayes dan KNN dapat memberikan hasil yang benar dan relevan sesuai dengan data yang tersedia.