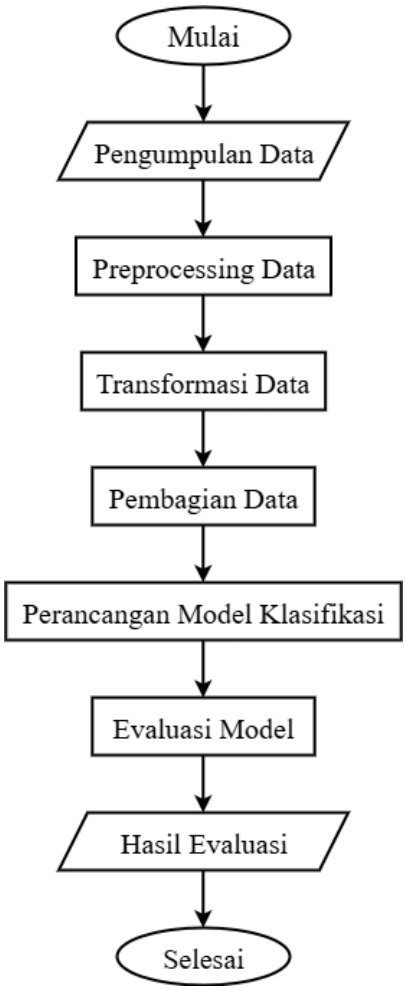
BAB II
LANDASAN TEORI

2.1. Pendahuluan

Pada bab ini, akan dibahas teori-teori yang menjadi dasar penelitian, termasuk analisis kepuasan, algoritma *Naive Bayes*, dan pengolahan bahasa alami (Natural Language Processing/NLP). Pendekatan ini digunakan untuk menganalisis opini masyarakat terhadap layanan rehabilitasi sosial di Rantauprapat. Analisis kepuasan telah menjadi alat penting untuk memahami persepsi masyarakat, khususnya dalam mengevaluasi kualitas layanan publik. Melalui analisis kepuasan, pengambil kebijakan dapat memperoleh wawasan berharga mengenai opini masyarakat, yang dapat menjadi dasar untuk meningkatkan layanan publik secara strategis.

Sebagai contoh, penelitian [1] menunjukkan bahwa analisis kepuasan pada aplikasi layanan publik seperti MySAPK BKN mengungkapkan berbagai keluhan masyarakat yang relevan. Hasil ini menjadi acuan penting untuk mengidentifikasi area perbaikan dalam pengembangan aplikasi tersebut. Selain itu, penelitian lain menunjukkan bahwa analisis kepuasan terhadap program vaksinasi COVID-19 di Indonesia memberikan informasi real-time tentang opini masyarakat. Temuan ini membantu dalam perencanaan program kesehatan yang lebih efektif dengan mengidentifikasi kepuasan yang berkembang di masyarakat [2]Lebih jauh lagi, analisis kepuasan memungkinkan pemerintah untuk membangun komunikasi yang lebih baik dengan masyarakat. Dengan menggunakan algoritma dan teknik analisis yang canggih, seperti pembelajaran mesin (machine learning), analisis ini mampu mendeteksi emosi dan sikap masyarakat secara lebih akurat. Hal ini mendukung

pembuatan kebijakan yang lebih responsif terhadap kebutuhan publik, sebagaimana diuraikan [3]Oleh karena itu, analisis kepuasan tidak hanya berfungsi sebagai alat evaluasi, tetapi juga sebagai sarana strategis untuk memperkuat hubungan antara pemerintah dan masyarakat.



Gambar 2. 1. Kerangka Penelitian

Gambar 2.1 Kerangka Penelitian di atas menunjukkan tahapan utama dalam proses analisis kepuasan, yang dimulai dari Menentukan Topik dan diakhiri dengan tahap Selesai. Setiap tahap memiliki peran penting dalam memastikan keberhasilan proses analisis kepuasan. Untuk penjelasan setiap tahapan yaitu sebagai berikut.

1. Pengumpulan data dilakukan dengan menyebarkan kuesioner digital kepada masyarakat yang pernah menerima layanan rehabilitasi sosial dari Dinas

Sosial Rantauprapat. Data yang dikumpulkan mencakup penilaian terhadap aspek layanan serta opini terbuka dari responden.

2. Preprocessing data dilakukan dengan membersihkan data dari entri yang kosong atau tidak relevan dan menyatukan skor dari setiap indikator layanan. Data kemudian diseleksi agar hanya mencakup variabel yang diperlukan dalam proses klasifikasi.
3. Transformasi data dilakukan dengan mengubah nilai numerik hasil kuesioner menjadi kategori seperti “Baik” dan “Tidak Baik”. Proses ini bertujuan agar data dapat diproses secara optimal oleh algoritma klasifikasi yang digunakan.
4. Pembagian data dilakukan dengan memisahkan data menjadi dua bagian, yaitu data training dan data testing. Pemisahan ini penting untuk memastikan bahwa model yang dibangun dapat diuji menggunakan data yang belum pernah dikenali sebelumnya.
5. Perancangan model klasifikasi dilakukan dengan menggunakan algoritma *Naive Bayes* yang mengandalkan prinsip probabilistik. Model ini dirancang untuk mengklasifikasikan opini masyarakat berdasarkan pola dari atribut layanan.
6. Evaluasi model dilakukan untuk menilai kinerja klasifikasi yang telah dibangun dalam mengenali kategori kepuasan masyarakat. Teknik evaluasi dilakukan melalui pengukuran tingkat kecocokan antara hasil prediksi dengan data aktual.

7. Hasil evaluasi digunakan sebagai dasar untuk menilai sejauh mana model mampu menjalankan tugas klasifikasinya dengan baik. Tahapan ini menjadi bagian penting dalam memastikan validitas dan keandalan pendekatan yang digunakan dalam penelitian.

2.2. Analisis Kepuasan

2.2.1. Pengertian Analisis Kepuasan

Analisis kepuasan merupakan suatu proses evaluatif yang bertujuan untuk mengukur sejauh mana harapan atau ekspektasi individu—dalam hal ini masyarakat—terpenuhi oleh suatu layanan, produk, atau interaksi yang diberikan oleh pihak penyelenggara layanan. Dalam konteks pelayanan publik, analisis kepuasan menjadi alat penting untuk memahami persepsi, penilaian, serta tanggapan masyarakat terhadap kualitas pelayanan yang mereka terima. Proses ini tidak hanya mencakup pengumpulan data melalui survei atau kuesioner, tetapi juga mencakup tahap interpretasi data guna menghasilkan informasi yang dapat digunakan untuk perbaikan berkelanjutan. Analisis kepuasan berfungsi sebagai jembatan antara penyedia layanan dan pengguna, di mana hasil dari analisis tersebut dapat menjadi dasar dalam pengambilan keputusan strategis, evaluasi kinerja, serta penyusunan kebijakan pelayanan yang lebih responsif dan berorientasi pada kebutuhan masyarakat. Dengan demikian, analisis kepuasan bukan hanya sekadar alat ukur, melainkan menjadi komponen penting dalam mewujudkan layanan publik yang efektif, adil, dan berkelanjutan.

2.2.2. Tahapan Analisis Kepuasan

Analisis kepuasan melibatkan beberapa tahapan utama yang dilakukan secara sistematis untuk memastikan keakuratan hasil. Tahapan-tahapan tersebut meliputi:

1. Pengumpulan data dilakukan dengan menyebarkan kuesioner digital kepada masyarakat yang pernah menerima layanan rehabilitasi sosial dari Dinas Sosial Rantauprapat. Data yang dikumpulkan mencakup penilaian terhadap aspek layanan serta opini terbuka dari responden.
2. Preprocessing data dilakukan dengan membersihkan data dari entri yang kosong atau tidak relevan dan menyatukan skor dari setiap indikator layanan. Data kemudian diseleksi agar hanya mencakup variabel yang diperlukan dalam proses klasifikasi.
3. Transformasi data dilakukan dengan mengubah nilai numerik hasil kuesioner menjadi kategori seperti “Baik” dan “Tidak Baik”. Proses ini bertujuan agar data dapat diproses secara optimal oleh algoritma klasifikasi yang digunakan.
4. Pembagian data dilakukan dengan memisahkan data menjadi dua bagian, yaitu data training dan data testing. Pemisahan ini penting untuk memastikan bahwa model yang dibangun dapat diuji menggunakan data yang belum pernah dikenali sebelumnya.
5. Perancangan model klasifikasi dilakukan dengan menggunakan algoritma *Naive Bayes* yang mengandalkan prinsip probabilistik. Model ini dirancang untuk mengklasifikasikan opini masyarakat berdasarkan pola dari atribut layanan.

6. Evaluasi model dilakukan untuk menilai kinerja klasifikasi yang telah dibangun dalam mengenali kategori kepuasan masyarakat. Teknik evaluasi dilakukan melalui pengukuran tingkat kecocokan antara hasil prediksi dengan data aktual.
7. Hasil evaluasi digunakan sebagai dasar untuk menilai sejauh mana model mampu menjalankan tugas klasifikasinya dengan baik. Tahapan ini menjadi bagian penting dalam memastikan validitas dan keandalan pendekatan yang digunakan dalam penelitian.

2.3. Algoritma Naive Bayes

Algoritma *Naive Bayes* adalah metode klasifikasi berbasis probabilitas yang menggunakan prinsip dasar Teorema Bayes. Metode ini mengasumsikan independensi antar fitur dalam dataset, meskipun asumsi ini jarang sepenuhnya terpenuhi dalam data dunia nyata. Meskipun demikian, *Naive Bayes* sering digunakan karena efisiensinya dalam mengolah data besar dan performanya yang kompetitif, terutama dalam analisis data teks. Dalam analisis kepuasan, algoritma ini diterapkan untuk memprediksi kategori kepuasan berdasarkan distribusi kata dalam dokumen.

2.3.1. Prinsip Dasar Teorema Bayes

Teorema Bayes yang menjadi dasar algoritma ini dirumuskan sebagai berikut:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

Sumber) [4][5]

Keterangan:

- P(C|X): Probabilitas bahwa dokumen X termasuk dalam kategori C (posterior).
- P(X|C): Probabilitas kemunculan fitur X dalam kategori C.
- P(C): Probabilitas prior kategori C secara keseluruhan.
- P(X): Probabilitas kemunculan fitur X di semua kategori.

Dalam konteks analisis kepuasan, fitur X seringkali direpresentasikan sebagai kata-kata dalam teks, sedangkan kategori C merepresentasikan kepuasan.

2.3.2. Keunggulan dan Kelemahan Algoritma *Naive Bayes*

1. Keunggulan

1. Sederhana dan Cepat: Algoritma ini cepat dalam melatih model dan membuat prediksi, menjadikannya pilihan ideal untuk aplikasi dengan data besar atau pemrosesan waktu nyata
2. Efisiensi pada Data Dimensi Tinggi: *Naive Bayes* mampu mengolah data dengan banyak fitur tanpa memerlukan banyak sumber daya komputasi [6]
3. Kinerja Baik pada Dataset Kecil: Algoritma ini tetap menunjukkan hasil yang baik meskipun menggunakan dataset yang terbatas, selama data memiliki distribusi representatif [7]

2. Kelemahan

1. Asumsi Independensi Fitur: Asumsi ini sering tidak realistis dalam data dunia nyata, terutama dalam analisis teks di mana kata-kata memiliki keterkaitan semantik [8]
2. Kurang Efektif pada Dataset Tidak Seimbang: Algoritma ini cenderung memberikan hasil bias terhadap kelas yang dominan

[9]Ketidakmampuan Menangkap Hubungan Non-Linear: *Naive Bayes* tidak dapat memodelkan hubungan kompleks antar fitur, yang dapat membatasi akurasi pada dataset yang memiliki pola non-linear [10]

2.3.3. *Implementasi dalam Analisis Kepuasan*

Dalam penelitian ini, algoritma *Naive Bayes* digunakan untuk menganalisis kepuasan ulasan pengguna. Kerangka kerja implementasi mencakup tahapan sebagai berikut:

1. Pengumpulan Data: Data ulasan pengguna dikumpulkan dari sumber digital seperti media sosial atau platform ulasan.
2. Preprocessing Data: Data diproses untuk menghilangkan elemen tidak relevan, melakukan tokenisasi, dan mentransformasikan teks menjadi vektor fitur menggunakan pendekatan Term Frequency-Inverse Document Frequency (TF-IDF).
3. Pelatihan Model: Algoritma *Naive Bayes* dilatih menggunakan data latih yang telah dikategorikan sesuai dengan kepuasan yang dimiliki.
4. Evaluasi Kinerja: Model dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score untuk menilai performa model dalam memprediksi kepuasan pada data uji.

Evaluasi model bertujuan untuk memastikan keandalan prediksi algoritma. Berikut adalah metrik evaluasi yang digunakan:

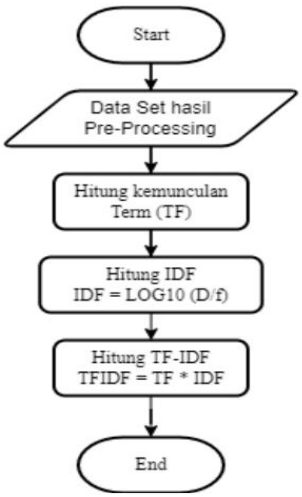
1. Akurasi: Proporsi prediksi yang benar dibandingkan total prediksi.
2. Precision: Proporsi prediksi positif yang benar-benar positif.

3. Recall: Proporsi data positif yang terdeteksi dengan benar oleh model.
4. F1-Score: Rata-rata harmonis antara precision dan recall, memberikan gambaran komprehensif tentang kinerja model.
5. Tabel dan Gambar yang Direkomendasikan
6. Tabel: Contoh Distribusi Data TF-IDF

Tabel dapat menampilkan contoh distribusi bobot kata berdasarkan metode TF-IDF untuk masing-masing kategori kepuasan.

Diagram: Alur Kerja Implementasi *Naive Bayes* dalam Analisis Kepuasan

Diagram ini dapat menunjukkan tahapan mulai dari pengumpulan data hingga evaluasi model untuk memberikan pemahaman menyeluruh tentang alur implementasi.



Gambar 2. 2. Diagram Alir Pembobotan TF-IDF

Diagram alir yang disajikan menggambarkan proses perhitungan Term Frequency-Inverse Document Frequency (TF-IDF) yang merupakan langkah penting dalam pemilihan fitur pada algoritma *Naive Bayes* untuk analisis kepuasan. Berikut adalah penjelasan setiap langkah dalam diagram:

1. Start: Proses dimulai setelah data teks selesai diproses dalam tahap preprocessing (misalnya tokenisasi, stopword removal, dan stemming).
2. Hitung Kemunculan Term (TF): Di sini, setiap kata (term) dalam dokumen dihitung frekuensinya, yaitu berapa kali kata tersebut muncul dalam teks.
3. Hitung IDF: IDF dihitung dengan rumus:

   $\mathrm{IDF} = \log_{10}\left(\frac{D}{f}\right)$, di mana D adalah jumlah total dokumen dan fff adalah jumlah dokumen yang mengandung kata tersebut. IDF bertujuan untuk memberikan bobot lebih pada kata-kata yang lebih jarang muncul di seluruh dokumen.
4. Hitung TF-IDF: Nilai TF-IDF untuk setiap kata dihitung dengan mengalikan nilai TF dan IDF. Hasilnya memberikan bobot yang menggambarkan pentingnya kata tersebut dalam konteks dokumen relatif terhadap keseluruhan dataset.
5. End: Proses perhitungan selesai dan fitur TF-IDF siap digunakan untuk pelatihan model *Naive Bayes*.

2.4. Penelitian Terdahulu

Penelitian terdahulu merupakan bagian penting yang menjadi dasar pijakan dan pembanding bagi penelitian yang sedang dilakukan. Melalui telaah pustaka ini, dapat diketahui bagaimana pendekatan yang telah digunakan oleh peneliti sebelumnya, serta kelebihan dan kekurangan dari masing-masing penelitian. Tinjauan ini juga berfungsi untuk mengidentifikasi celah (gap) penelitian yang belum banyak dibahas, sehingga penelitian ini dapat memberikan kontribusi ilmiah yang lebih signifikan.

Berikut adalah beberapa penelitian terdahulu yang relevan dengan topik analisis kepuasan dan pemanfaatan algoritma *Naive Bayes*, terutama dalam konteks layanan publik dan opini masyarakat:

Tabel 2. 1. Penelitian Terdahulu

Referensi Penelitian	1
Judul	Comparison Of The C.45 And Naive Bayes Algorithms To Predict Diabetes
Nama	Alam1)*, Divi Adiffia Freza Alana.2), Christina Juliane3)
Tahun	2023
Hasil	Metode **Naive Bayes** telah banyak digunakan secara efektif dalam berbagai penelitian, termasuk pada kasus deteksi diabetes, dengan tingkat akurasi yang dilaporkan mencapai sekitar 90%. Algoritma ini bekerja dengan prinsip probabilitas dan asumsi "naive" bahwa setiap fitur atau atribut bersifat independen satu sama lain. Keunggulan Naive Bayes terletak pada kemampuannya dalam mengolah data medis dengan sederhana namun tetap memberikan hasil klasifikasi yang akurat. Oleh karena itu, penerapan metode Naive Bayes dalam konteks penelitian sebelumnya menunjukkan bahwa algoritma ini berpotensi besar dalam mendukung proses identifikasi penyakit serta menjadi dasar

	pengembangan sistem klasifikasi kesehatan yang lebih efektif [11].
Referensi Penelitian	2
Judul	Performance of Various Naïve Bayes Using GridSearch Approach In Phishing Email Dataset
Nama	Rizki Rahman1)*, Ferian Fauzi Abdulloh2)
Tahun	2023
Hasil	Penerapan metode Naive Bayes dalam klasifikasi email phishing terbukti sangat efektif, dengan tingkat akurasi mencapai sekitar 97%. Metode ini bekerja dengan prinsip probabilistik sederhana yang mengasumsikan independensi antar fitur, namun tetap mampu memberikan performa tinggi dalam mendeteksi pola pada data teks. Pada penelitian ini, data email phishing melalui tahap pra-pemrosesan yang mencakup pembersihan teks, transformasi ke bentuk numerik, serta ekstraksi fitur yang relevan untuk dianalisis. Evaluasi dilakukan terhadap beberapa varian Naive Bayes, yang berbeda dalam cara menghitung probabilitas dan penanganan distribusi data, namun hasilnya menunjukkan konsistensi kinerja yang unggul. Temuan ini menegaskan bahwa metode

	Naive Bayes, dengan dukungan tahapan pra-pemrosesan yang tepat, dapat menjadi pendekatan yang efisien dan handal dalam mengidentifikasi email phishing [12].
Referensi Penelitian	3
Judul	Analisis Layanan Pelanggan PT PLN Berdasarkan Media Sosial Twitter Dengan Menggunakan Metode Naïve Bayes Classifier
Nama Penulis	Tri Prasetyo[1], Hadi Zakaria[2*], Pandu Wiliantoro[3*]
Tahun	2022
Hasil	Metode **Naive Bayes** telah banyak diterapkan sebagai salah satu teknik klasifikasi yang efektif dalam menganalisis data kepuasan masyarakat terhadap layanan PT. PLN. Dengan pendekatan probabilistik yang sederhana namun kuat, algoritma ini mampu mengolah data umpan balik dan evaluasi masyarakat untuk menemukan pola yang relevan dalam menentukan tingkat kepuasan.

	Penelitian terdahulu menunjukkan bahwa penerapan metode Naive Bayes menghasilkan tingkat akurasi sebesar 80%, yang membuktikan bahwa model ini cukup handal dalam mengklasifikasikan kategori kepuasan masyarakat berdasarkan data yang tersedia. Hal ini menegaskan bahwa Naive Bayes dapat menjadi metode yang tepat dalam mendukung proses analisis data evaluasi layanan [13].
Referensi Penelitian	4
Judul	Sistem Pakar Diagnosa Gangguan Jiwa Menggunakan Metode Naive Bayes Berbasis
Nama Penulis	Meriyam Yunita[1)], Tri Widodo[2)]
Tahun	2021
Hasil	Metode Naive Bayes merupakan salah satu teknik klasifikasi berbasis probabilitas yang banyak digunakan dalam penelitian, termasuk untuk analisis gejala gangguan jiwa seperti skizofrenia paranoid. Algoritma ini

	bekerja dengan menerapkan teorema Bayes untuk menghitung peluang suatu data masuk ke dalam kategori tertentu berdasarkan fitur yang dimiliki. Meskipun Naive Bayes mengasumsikan bahwa setiap fitur bersifat independen, metode ini terbukti sederhana, efisien, dan cukup akurat dalam mengenali pola gejala yang muncul. Oleh karena itu, Naive Bayes sering dimanfaatkan dalam penelitian terdahulu untuk membantu proses klasifikasi gejala skizofrenia paranoid secara sistematis dan berbasis data [14].
Referensi Penelitian	5
Judul	Prediksi Tingkat Kepuasan dalam Pembelajaran Daring Menggunakan Algoritma Naïve Bayes
Nama Penulis	Abdi Rahim Damanik1 ,Sumijan2, Gunadi Widi Nurcahyo3
Tahun	2021

Hasil	Penelitian ini menggunakan metode **Naïve Bayes** untuk memprediksi tingkat kepuasan mahasiswa dalam pembelajaran daring pada masa pandemi COVID-19. Data diperoleh dari kuesioner yang mencakup indikator komunikasi dosen, suasana pembelajaran, penilaian mahasiswa, dan penyampaian materi. Hasil pengujian dengan RapidMiner menggunakan data training dan testing menunjukkan tingkat akurasi, precision, dan recall masing-masing sebesar **100%**, sehingga metode Naïve Bayes dapat direkomendasikan sebagai model prediksi yang efektif dalam mengukur kepuasan mahasiswa terhadap pembelajaran online [15].
Referensi Penelitian	6
Judul	Komparasi Model Klasifikasi Sentimen Issue Vaksin Covid-19 Berbasis Platform Instagram

Nama Penulis	Primandani Arsi* , Laili Nur Hidayati, Azizan Nurhakim
Tahun	2022
Hasil	Topik vaksinasi covid-19 di Instagram, khususnya pada akun resmi **kemenkes_ri**, menimbulkan beragam opini masyarakat yang terbagi dalam sentimen positif, negatif, dan netral. Penelitian ini berfokus pada analisis sentimen menggunakan metode **Naïve Bayes** dengan data sebanyak 2.925 komentar hasil scrapping. Tahapan penelitian meliputi preprocessing teks dengan teknik NLP, pembobotan kata menggunakan TF-IDF, serta penanganan ketidakseimbangan data dengan SMOTE, kemudian dilakukan pembagian data training dan testing dengan rasio 9:1. Hasil klasifikasi menunjukkan bahwa algoritma Naïve Bayes mampu memberikan performa cukup baik dengan akurasi sebesar **87,65%**, presisi **85,40%**, dan recall

	86,20%, sehingga dapat disimpulkan bahwa metode ini efektif digunakan dalam mengklasifikasikan opini publik terkait kebijakan vaksinasi di media sosial [16].
Referensi Penelitian	7
Judul	PENERAPAN METODE NAÏVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE PADA ANALISIS SENTIMEN TERHADAP DAMPAK VIRUS CORONA DI TWITTER
Nama Penulis	Cholid Fadilah Hasri1, Debby Alita2
Tahun	2022
Hasil	Penggunaan media sosial, khususnya Twitter, memungkinkan masyarakat menyebarkan dan memperoleh informasi dengan cepat, termasuk opini terkait kebijakan PPKM di Indonesia sepanjang 2021. Penelitian terdahulu memanfaatkan metode klasifikasi Naïve Bayes untuk menganalisis sentimen masyarakat terhadap

	kebijakan tersebut, dengan tujuan mengklasifikasikan opini menjadi sentimen positif, netral, atau negatif. Hasil penelitian menunjukkan bahwa Naïve Bayes Classifier mampu mengklasifikasi data dengan baik, menghasilkan akurasi sebesar 81,07%, presisi 80,12%, dan recall 79,85%, sehingga metode ini terbukti efektif dalam menganalisis sentimen opini masyarakat di Twitter tanpa memerlukan penambahan fitur tambahan [17].
Referensi Penelitian	8
Judul	PENERAPAN METODE NAÏVE BAYES DAN C4.5 PADA PENERIMAAN PEGAWAI DI UNIVERSITAS POTENSI UTAMA
Nama Penulis	Cindy Paramitha Lubis1, Dr. Zakarias Situmorang, MT2
Tahun	2020
Hasil	Penelitian terdahulu fokus pada penerapan metode **Naïve Bayes** untuk

	seleksi dan klasifikasi calon pegawai yang potensial masuk ke kampus dengan menghitung probabilitas pada setiap kriteria yang ada. Permasalahan yang sering muncul adalah ketidakefektifan metode dalam menyesuaikan calon pegawai dengan bidang keahlian yang dibutuhkan. Pengujian dilakukan menggunakan tools **Weka 3.8** pada 36 data latih, menghasilkan tingkat akurasi **77,78%**, dengan presisi dan recall yang menunjukkan kemampuan model dalam mengklasifikasikan data secara tepat. Hasil ini menegaskan bahwa metode Naïve Bayes mampu memberikan klasifikasi yang cukup baik, meskipun masih ada ruang untuk peningkatan dalam menyesuaikan kandidat dengan kebutuhan spesifik institusi [18].
Referensi Penelitian	9

Judul	Implementation of the Naïve Bayes Method to determine the Level of Consumer Satisfaction
Nama Penulis	Fitri Febriyani Hasibuan1)*, Muhammad Halmi Dar2), Gomal Juni Yanris3)
Tahun	2023
Hasil	Penelitian ini bertujuan untuk menganalisis tingkat kepuasan konsumen di Brastagi Supermarket dengan menggunakan metode Naïve Bayes. Data yang digunakan berjumlah 49 konsumen, kemudian diproses melalui tahapan analisis data, preprocessing, penerapan algoritma Naïve Bayes, hingga pengujian sistem menggunakan aplikasi Orange. Hasil klasifikasi menunjukkan bahwa 47 konsumen puas dan 2 konsumen tidak puas terhadap pengalaman belanja mereka. Evaluasi model menggunakan widget *test and score* menghasilkan akurasi 92,9%, sedangkan evaluasi

	dengan *confusion matrix* memberikan akurasi 97,9%, menunjukkan presisi dan recall yang tinggi. Dengan demikian, metode Naïve Bayes terbukti efektif dalam memprediksi kepuasan konsumen, menghasilkan klasifikasi yang akurat dan dapat diandalkan untuk menilai kualitas layanan serta produk yang diberikan [19].
Referensi Penelitian	10
Judul	Implementation of Data Mining for Data Classification of Visitor Satisfaction Levels
Nama Penulis	Hubban Arfi Pratama1)*, Gomal Juni Yanris2), Mila Nirmala Sari Hasibuan3)
Tahun	2023
Hasil	Berdasarkan penelitian sebelumnya, metode Naïve Bayes digunakan untuk mengklasifikasikan tingkat kepuasan pengunjung Boombara Waterpark Rantauprapat. Dari 100 data pengunjung, hasil klasifikasi

	menunjukkan bahwa 77 pengunjung merasa puas dan 23 pengunjung tidak puas. Evaluasi model menggunakan Confusion Matrix menghasilkan akurasi sebesar 99%, dengan nilai presisi dan recall yang juga tinggi, menunjukkan bahwa metode Naïve Bayes sangat efektif dalam memprediksi kepuasan pengunjung. Hasil ini menegaskan bahwa metode Naïve Bayes dapat dijadikan referensi yang handal untuk klasifikasi data kepuasan pengunjung di amusement park [20].

2.5. Keunggulan Penelitian Ini Dibandingkan Penelitian Terdahulu

Berdasarkan kajian teori adapun yang menjadi penelitian terdahulu sesuai dengan penelitian.

Tabel 2. 2. Perbandingan Metode dalam Penelitian Terdahulu dan Penelitian ini

Aspek Penelitian	Penelitian Terdahulu	Penelitian Ini
Fokus Penelitian	Analisis kepuasan pada kebijakan dan produk nasional	Analisis kepuasan dalam pelayanan sosial lokal
Metode Analisis Kepuasan	*Naive Bayes*, VADER, SVM, Lexicon-Based	*Naive Bayes* dengan TF-IDF untuk teks berbahasa Indonesia
Konteks	Kota besar atau nasional	Daerah lokal (Rantauprapat)

Pendekatan Preprocessing	Standar preprocessing (stopword removal, tokenization)	Penggunaan TF-IDF untuk fitur yang lebih relevan
Keunggulan	Akurasi tinggi pada dataset besar dan terstandarisasi	Menyesuaikan konteks lokal dan analisis bahasa Indonesia
Keterbatasan	Tidak selalu relevan untuk konteks lokal	Membutuhkan pemrosesan yang lebih teliti terhadap bahasa lokal

Penelitian ini memiliki keunggulan signifikan dibandingkan dengan penelitian terdahulu, khususnya dalam hal fokus pada analisis kepuasan dalam sektor pelayanan sosial di daerah lokal, seperti Rantauprapat. Penelitian terdahulu lebih banyak berfokus pada analisis kepuasan di kawasan urban besar yang memiliki karakteristik demografis dan sosial yang berbeda. Hal ini menciptakan celah dalam pemahaman tentang bagaimana kepuasan publik diekspresikan dalam konteks pelayanan sosial di daerah yang lebih kecil dan dengan karakteristik masyarakat yang berbeda. Dengan demikian, penelitian ini menawarkan perspektif baru yang lebih relevan dan aplikatif, terutama dalam meningkatkan

interaksi antara pemerintah dan masyarakat lokal melalui pemahaman yang lebih mendalam terhadap kepuasan publik terhadap layanan sosial yang diberikan.

Salah satu keunggulan utama dalam penelitian ini adalah penerapan algoritma *Naive Bayes* dalam analisis kepuasan terhadap data berbahasa Indonesia. Sebagaimana diketahui, analisis kepuasan dalam bahasa Indonesia menghadirkan tantangan tersendiri, terkait dengan struktur bahasa, kosakata, dan nuansa budaya yang khas. Dalam konteks ini, *Naive Bayes* dipilih karena kesederhanaannya dan kemampuannya untuk menangani data teks dalam jumlah besar, meskipun dengan asumsi independensi antar fitur yang ada. Penelitian ini berupaya mengatasi

tantangan tersebut dengan menerapkan preprocessing yang teliti dan penggunaan teknik Term Frequency-Inverse Document Frequency (TF-IDF) dalam pemrosesan data teks. Teknik ini memungkinkan untuk mengidentifikasi kata-kata yang memiliki bobot lebih penting dalam teks, yang pada gilirannya dapat meningkatkan akurasi model dalam mengklasifikasikan kepuasan. Dengan kata lain, penggunaan TF-IDF membantu untuk memfokuskan model pada kata-kata kunci yang relevan, sehingga model dapat memberikan hasil yang lebih tepat dan signifikan.

Sebagai perbandingan, penelitian sebelumnya lebih banyak menggunakan pendekatan yang mengandalkan metode *Naive Bayes* tanpa mempertimbangkan elemen-elemen linguistik khas dalam analisis bahasa Indonesia, seperti yang ditemukan pada penelitian [21] dan [22] yang berfokus pada kebijakan atau produk di tingkat nasional. Meskipun metode tersebut menunjukkan hasil yang memadai, namun belum sepenuhnya mengakomodasi perbedaan kontekstual dan sosial yang ada di daerah lokal. Selain itu, beberapa penelitian sebelumnya menggunakan dataset yang lebih besar dan lebih terstandarisasi, seperti yang terlihat pada penelitian [23], yang lebih mengutamakan klasifikasi kepuasan pada media sosial terkait kebijakan publik, namun tidak sepenuhnya menyesuaikan dengan kondisi dan konteks lokal.

Di sisi lain, penelitian ini menawarkan potensi untuk memberikan dampak yang lebih langsung dalam meningkatkan kualitas pelayanan publik di Rantauprapat. Dengan memahami bagaimana masyarakat lokal mengekspresikan kepuasan terhadap layanan publik, penelitian ini memberikan wawasan yang dapat digunakan untuk merancang kebijakan yang lebih responsif terhadap kebutuhan dan

harapan masyarakat. Hal ini akan berkontribusi pada peningkatan interaksi antara pemerintah dan masyarakat, serta memperbaiki efektivitas pelayanan sosial yang diberikan.

Dengan demikian, penelitian ini tidak hanya memperkaya literatur terkait analisis kepuasan dalam sektor pelayanan sosial, tetapi juga menawarkan solusi praktis dan aplikatif untuk meningkatkan kualitas layanan di daerah dengan karakteristik lokal yang lebih spesifik. Pendekatan yang lebih terfokus pada konteks lokal, serta penerapan teknik-teknik preprocessing dan seleksi fitur yang lebih relevan, menjadikan penelitian ini lebih inovatif dan signifikan dibandingkan dengan penelitian terdahulu.

2.6. Model Klasifikasi

2.6.1. Pengumpulan Data

Langkah pertama dalam kerangka kerja penelitian ini adalah pengumpulan data. Data diperoleh melalui penyebaran kuesioner digital yang dirancang secara khusus untuk menjaring opini masyarakat mengenai layanan rehabilitasi sosial yang diberikan oleh Dinas Sosial Rantauprapat. Kuesioner tersebut dibagikan secara daring menggunakan platform Google Form dan disebarluaskan melalui berbagai saluran komunikasi masyarakat lokal, seperti media sosial, grup komunitas, dan jaringan perangkat desa. Pertanyaan dalam kuesioner mencakup atribut seperti *Kualitas Pelayanan*, *Aksesibilitas Layanan*, *Transparansi Informasi*, serta kolom isian untuk menilai *Kepuasan Masyarakat*. Dengan menggunakan pendekatan ini, data yang terkumpul menjadi lebih terarah, relevan, dan sesuai dengan tujuan penelitian dalam mengklasifikasikan tingkat kepuasan masyarakat berdasarkan tanggapan mereka terhadap layanan yang telah diterima.

2.6.2. Preprocessing Data

Preprocessing data merupakan tahap penting dalam pengolahan data sebelum dilakukan analisis lebih lanjut atau pemodelan klasifikasi. Dalam penelitian ini, preprocessing dilakukan terhadap data hasil kuesioner yang telah dikumpulkan untuk memastikan bahwa data bersih, konsisten, dan sesuai dengan format yang dibutuhkan oleh algoritma *Naive Bayes*. Proses ini meliputi pemeriksaan kelengkapan data, penghapusan entri yang kosong atau tidak relevan, serta standarisasi nilai atribut menjadi bentuk kategorikal seperti "Baik" dan "Tidak Baik" untuk variabel layanan, serta "Puas" dan "Tidak Puas" untuk variabel target. Tahap ini bertujuan agar data lebih terstruktur dan dapat dengan mudah diproses oleh sistem klasifikasi, sekaligus meminimalkan potensi kesalahan yang dapat memengaruhi akurasi model yang dibangun.

2.6.3. Ekstraksi Fitur

Ekstraksi fitur adalah proses penting dalam pembuatan model *Naive Bayes*. Dalam penelitian ini, teknik TF-IDF digunakan untuk menentukan kata-kata yang memiliki bobot penting dalam teks dan untuk membantu model dalam melakukan klasifikasi yang lebih akurat [24] Ekstraksi fitur ini penting karena pemilihan kata-kata yang relevan dapat meningkatkan performa model. Teknik lain seperti Unigram, Bigram, atau N-Gram juga dapat digunakan untuk meningkatkan hasil ekstraksi fitur [25]

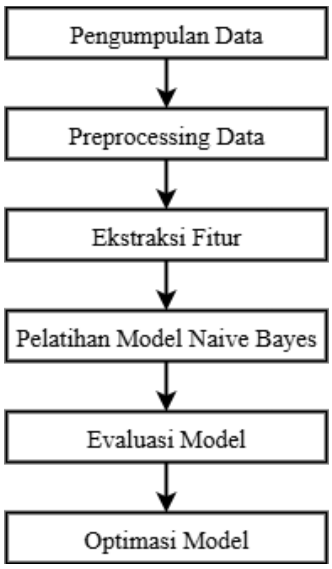
2.6.4. Pelatihan dan Evaluasi Model

Setelah fitur diekstraksi, model *Naive Bayes* dilatih menggunakan dataset yang telah diproses. Evaluasi model dilakukan menggunakan metrik seperti presisi,

recall, dan F1-score, yang memberikan gambaran lebih komprehensif tentang kinerja model [26] Evaluasi ini penting untuk mengukur sejauh mana model mampu mengklasifikasikan kepuasan secara akurat, terutama dalam hal mengidentifikasi kepuasan negatif, yang sering kali lebih sulit dianalisis [27]

2.6.5. Optimasi Model

Untuk meningkatkan akurasi model, beberapa penelitian telah mengusulkan penggunaan teknik optimasi. Salah satunya adalah Particle Swarm Optimization (PSO), yang dapat membantu dalam menyesuaikan parameter model *Naive Bayes* untuk mencapai performa yang lebih baik [28]. Berikut adalah representasi flowchart yang lebih sederhana dan jelas:



Gambar 2. 3. Flowchart Kerangka Kerja Penelitian

Penjelasan Ringkas Setiap Langkah:

1. Pengumpulan Data

   Pada tahap ini, kita mengumpulkan data opini masyarakat dari berbagai platform digital. Data yang dikumpulkan biasanya berupa teks yang

mencakup opini, komentar, atau ulasan terkait layanan sosial. Hal ini dilakukan untuk memperoleh *feedback* langsung dari masyarakat.

2. Preprocessing Data

   Setelah data terkumpul, langkah berikutnya adalah membersihkan data dari elemen yang tidak diperlukan seperti simbol, angka, atau kata-kata umum yang tidak memiliki makna. Proses ini memungkinkan model untuk bekerja dengan data yang lebih bersih dan relevan.

3. Ekstraksi Fitur

   Di tahap ini, data teks yang telah dibersihkan akan diubah menjadi data numerik. TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah metode yang digunakan untuk mengukur pentingnya suatu kata dalam dokumen relatif terhadap seluruh korpus. Ini memungkinkan model untuk menghitung relevansi setiap kata dalam analisis kepuasan.

4. Pelatihan Model *Naive Bayes*

   Pada tahap ini, model *Naive Bayes* dilatih untuk mempelajari hubungan antara fitur-fitur yang diekstraksi dan kategori kepuasan yang relevan. *Naive Bayes* mengandalkan probabilitas untuk mengklasifikasikan setiap opini ke dalam kelas yang tepat.

5. Evaluasi Model

   Evaluasi adalah tahap di mana model diuji dengan data uji yang tidak digunakan selama pelatihan untuk memastikan model tidak overfitting.

2.7. Tools Pendukung Penelitian

Dalam menunjang proses klasifikasi data sosial menggunakan algoritma *Naive Bayes*, penelitian ini menggunakan beberapa perangkat lunak dan tools pendukung untuk memastikan akurasi perhitungan, efisiensi pemrosesan, serta kemudahan dalam visualisasi dan analisis data. Adapun tools yang digunakan meliputi:

1. Microsoft Excel

Digunakan dalam tahap awal untuk proses input data, pengolahan awal, serta perhitungan manual seperti probabilitas dasar, distribusi frekuensi, dan pembuatan confusion matrix.

Gambar 2. 4. Aplikasi Microsoft Excel

2. Rapid Miner

Merupakan platform data science yang digunakan untuk membangun, melatih, dan mengevaluasi model klasifikasi *Naive Bayes*. RapidMiner

Gambar 2. 5. Aplikasi RapidMiner