BAB II

KAJIAN PUSTAKA

A. Kajian Teoritis

1. Sentimen Analisis

Sentimen analisis merupakan ilmu pengetahuan yang digunakan untuk mengetahui kecenderungan opini dari beberapa orang atau masyarakat, biasanya dalam bentuk komentar atau tulisan, di mana informasi berharga dapat ditemukan dari data yang tidak terstruktur dan diklasifikasikan ke dalam sentimen positif atau negatif berdasarkan emosional opini yang diberikan, dengan pengelompokan ini dilakukan menggunakan klasifikasi teks (Ramadhani & Suryono, 2024). Sentimen analisis adalah ilmu yang menganalisis data dengan memeriksa berbagai sudut pandang, sentimen, dan emosi yang disampaikan dalam teks untuk menentukan apakah nilainya positif atau negative, proses analisis sentimen ini dapat dilakukan menggunakan algoritma klasifikasi seperti Naive Bayes (Kurniawan et al., 2023).

Sentimen analisis adalah metode untuk mengidentifikasi dan mengklasifikasikan opini atau emosi dalam teks, seperti ulasan produk, komentar media sosial, atau feedback pelanggan, tujuannya adalah menentukan sikap penulis terhadap topik, produk, atau layanan, sehingga membantu bisnis memahami persepsi pelanggan dan membuat keputusan lebih tepat berdasarkan analisis data teks yang tidak terstruktur (Alam et al., 2024).

Dapat disimpulkan bahwa analisis sentimen adalah metode yang digunakan untuk mengetahui opini atau emosi seseorang dalam teks, seperti komentar, ulasan, atau feedback. Analisis ini mengklasifikasikan sentimen menjadi positif atau negatif dengan menggunakan algoritma seperti Naive Bayes. Dengan demikian, metode ini membantu bisnis atau organisasi memahami persepsi masyarakat dan mengambil keputusan lebih tepat berdasarkan informasi dari data tidak terstruktur.

2. Aduan Pelayanan Publik

Pelayanan publik adalah kegiatan yang bertujuan memenuhi kebutuhan masyarakat melalui penyediaan jasa, barang, dan layanan administratif. Semua kegiatan ini dilakukan sesuai peraturan perundang-undangan yang berlaku dan diselenggarakan oleh instansi atau lembaga penyedia pelayanan publik untuk menjamin hak dan kepuasan masyarakat secara adil dan merata (Bazarah, 2023). Aduan pelayanan publik adalah laporan atau keluhan yang disampaikan oleh masyarakat terkait dengan kualitas atau kinerja pelayanan yang diberikan oleh instansi pemerintah, yang bertujuan untuk memperbaiki atau meningkatkan layanan publik agar lebih efektif, efisien, dan memenuhi kebutuhan masyarakat (Akib, 2022).

Aduan pelayanan publik adalah komunikasi warga yang menyampaikan ketidakpuasan atau masalah dalam menerima layanan pemerintah, mencakup aspek prosedur, kualitas, dan kecepatan, tujuannya agar instansi terkait menindaklanjuti dan memperbaiki layanan, sehingga

meningkatkan kepuasan masyarakat terhadap pelayanan publik secara efektif (Theresia et al., 2025).

Dapat disimpulkan bahwa pelayanan publik merupakan kegiatan yang bertujuan memenuhi kebutuhan masyarakat melalui penyediaan jasa, barang, dan layanan administratif sesuai peraturan yang berlaku. Aduan pelayanan publik muncul sebagai bentuk laporan atau keluhan masyarakat terkait kualitas dan kinerja layanan, yang bertujuan memperbaiki efektivitas, efisiensi, serta kepuasan masyarakat agar layanan publik dapat berjalan lebih optimal.

3. Metode Naive Bayes

Metode Naive Bayes adalah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan, menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas (Martantoh et al., 2022), berikut rumus persamaan (2.1).

$$P(H|X) = \frac{P(X|Ci).P(Ci)}{P(X)} \quad (2.1)$$

Pada rumus Naive Bayes, X merupakan data yang kelasnya belum diketahui, sedangkan Ci merupakan hipotesis bahwa data tersebut termasuk ke dalam suatu kelas tertentu. Nilai P(Ci|X) menunjukkan probabilitas bahwa data X termasuk ke dalam kelas Ci berdasarkan kondisi atau fitur yang dimiliki data tersebut, sehingga disebut sebagai probabilitas

posterior. Nilai P(Ci) menunjukkan probabilitas awal suatu kelas sebelum data X dianalisis, sehingga disebut probabilitas prior. Selanjutnya, P(X|Ci) menunjukkan probabilitas kemunculan data X apabila diketahui data tersebut berada pada kelas Ci. Sementara itu, P(X) merupakan probabilitas kemunculan data X secara keseluruhan. Melalui perhitungan tersebut, algoritma Naive Bayes dapat menentukan kelas yang paling sesuai untuk suatu data berdasarkan nilai probabilitas tertinggi.

Teorema Bayes sering pula dikembangkan mengingat berlakunya probabilitas total, berikut rumus persamaan (2.2).

$$P(H|X) = \frac{P(X|H)}{\sum_{i=1}^{n} P(H_i|X)} \times P(H) \tag{2.2}$$

Teorema Bayes digunakan untuk menghitung probabilitas akhir suatu hipotesis Hsetelah diberikan bukti X. Nilai $P(H \mid X)$disebut sebagai probabilitas posterior, yaitu probabilitas setelah observasi data. Sedangkan $P(H)$merupakan probabilitas awal atau prior, dan $P(X \mid H)$adalah likelihood, yaitu peluang munculnya bukti Xjika hipotesis Hbenar. Seluruh hipotesis yang mungkin, yaitu $H_1, H_2, \ldots, H_n$, membentuk ruang sampel Ssehingga seluruh kemungkinan kejadian tercakup dalam sistem.

Untuk memahami Teorema Naïve Bayes, harus diperhatikan bahwa siklus klasifikasi memerlukan berbagai tanda untuk mengetahui kelas apa yang sesuai untuk contoh yang sedang diselidiki. berikut rumus persamaan (2.3).

$$P(C|F_{1,\dots}F_n) = \frac{P(F_1, \dots F_n|C)}{P(F_1, \dots F_n)} \times P(C) \quad (2.3)$$

Rumus di atas dapat pula ditulis secara sederhana sebagai berikut:

$$Posterior = \frac{prior\ x\ Likehood}{evidence} \quad (2.4)$$

Dengan demikian, klasifikasi berdasarkan Naive Bayes dapat dihitung dengan menghitung nilai probabilitas untuk setiap kelas dan memilih kelas dengan probabilitas tertinggi. Metode Naive Bayes adalah algoritma klasifikasi berbasis probabilitas yang menggunakan teorema Bayes untuk menghitung kemungkinan suatu data termasuk ke dalam kategori tertentu, dengan mengasumsikan bahwa setiap atribut dalam data tersebut bersifat independen satu sama lain, dan sering digunakan dalam pengolahan teks seperti analisis sentimen dan pemfilteran spam (Kumar & Goswami, 2024).

Metode Naive Bayes adalah teknik klasifikasi yang menggunakan probabilitas untuk memprediksi kategori data berdasarkan nilai atribut yang dimiliki, dengan mengasumsikan bahwa setiap atribut bersifat independen satu sama lain, dan biasanya diterapkan dalam aplikasi seperti deteksi spam, analisis sentimen, serta pengklasifikasian dokumen teks (Bisono & Zulherry, 2025). Metode ini efektif untuk memprediksi kategori data dalam berbagai aplikasi, seperti analisis sentimen, deteksi spam, dan pengklasifikasian dokumen, karena sederhana, cepat, dan menghasilkan prediksi yang akurat.

4. Pengumpulan Data

Pengumpulan data merupakan tahap pertama yang sangat penting dalam analisis sentimen, karena kualitas dan kuantitas data yang dikumpulkan akan mempengaruhi hasil akhir dari analisis tersebut, di mana pada tahap ini, data yang relevan dengan topik atau tujuan analisis dikumpulkan dari berbagai sumber, seperti ulasan produk, komentar di media sosial, atau forum diskusi, yang dapat diperoleh melalui metode otomatis seperti scraping atau penggunaan API, maupun secara manual (Singh & Chandra, 2023). Pengumpulan data adalah proses mengumpulkan informasi yang relevan dan diperlukan untuk tujuan analisis atau penelitian, yang dapat dilakukan melalui berbagai metode seperti survei, wawancara, pengamatan, atau teknik otomatis seperti scraping dan penggunaan API, dengan tujuan untuk memperoleh data yang akurat dan representatif untuk mendukung kesimpulan atau keputusan yang akan diambil (Saiful et al., 2025).

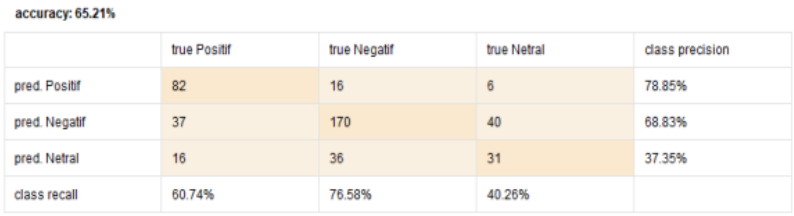
Pengumpulan data adalah langkah awal dalam penelitian atau analisis yang bertujuan untuk mengumpulkan informasi yang diperlukan dari berbagai sumber, baik itu secara langsung melalui pengamatan, wawancara, atau eksperimen, maupun secara tidak langsung menggunakan teknik seperti pengambilan data dari database, dokumen, atau sumber digital lainnya, yang kemudian akan digunakan untuk menganalisis suatu fenomena atau masalah (Rifa et al., 2023).

Dapat disimpulkan bahwa pengumpulan data merupakan tahap awal yang sangat penting dalam penelitian atau analisis, termasuk analisis sentimen, karena kualitas dan kuantitas data akan memengaruhi hasil akhir. Proses ini dilakukan dengan mengumpulkan informasi yang relevan dari berbagai sumber, baik secara langsung melalui wawancara, pengamatan, atau eksperimen, maupun secara tidak langsung melalui teknik otomatis seperti scraping atau penggunaan API, untuk memastikan data akurat dan representatif.

5. Preprocessing Teks

Preprocessing teks adalah tahap krusial dalam analisis sentimen yang bertujuan mempersiapkan data agar algoritma dapat memprosesnya secara efektif. Proses ini meliputi pembersihan teks, penghapusan karakter tidak relevan, normalisasi, tokenisasi, dan transformasi data menjadi format yang terstruktur sehingga mempermudah analisis dan meningkatkan akurasi model (Palomino, 2022). Preprocessing teks adalah tahap awal dalam pengolahan data teks yang bertujuan untuk membersihkan dan mempersiapkan teks agar dapat dianalisis lebih lanjut, dengan melakukan berbagai langkah seperti penghapusan tanda baca, konversi huruf kapital menjadi huruf kecil, penghilangan kata-kata tidak penting (stopwords), serta stemming atau lemmatization untuk mengubah kata ke bentuk dasarnya, sehingga teks menjadi lebih terstruktur dan siap untuk proses analisis seperti klasifikasi atau pemodelan (Elov et al., 2023).

Preprocessing teks adalah serangkaian langkah yang dilakukan untuk mempersiapkan data teks sebelum diproses lebih lanjut dalam analisis, seperti membersihkan teks dari karakter-karakter yang tidak relevan, menghilangkan kata-kata umum yang tidak memiliki makna penting (stopwords), mengonversi teks ke format yang lebih konsisten, serta melakukan normalisasi kata melalui stemming atau lemmatization, agar teks dapat lebih mudah diolah oleh algoritma dalam tugas seperti analisis sentimen atau klasifikasi (Hakim, 2021).



accuracy: 65.21%

	true Positif	true Negatif	true Netral	class precision
pred. Positif	82	16	6	78.85%
pred. Negatif	37	170	40	68.83%
pred. Netral	16	36	31	37.35%
class recall	60.74%	76.58%	40.26%	

Gambar 2. 1 Analasisi Preprocessing Teks
Sumber : (Sandy et al., 2024)

Dapat disimpulkan bahwa preprocessing teks merupakan tahap krusial dalam analisis sentimen yang bertujuan mempersiapkan data agar algoritma dapat memprosesnya secara efektif. Proses ini mencakup pembersihan teks, penghapusan karakter atau kata tidak relevan, normalisasi, tokenisasi, serta transformasi data menjadi format terstruktur, sehingga mempermudah analisis, meningkatkan akurasi model, dan memastikan data siap untuk proses klasifikasi atau pemodelan lebih lanjut.

6. Klasifikasi Sentimen

Klasifikasi sentimen adalah teknik analisis teks yang bertujuan mengidentifikasi dan mengelompokkan opini atau pendapat dalam suatu

teks, proses ini menentukan apakah teks mengandung sentimen positif, negatif, atau netral, sehingga memungkinkan pemahaman terhadap persepsi, opini, dan sikap penulis dalam konteks produk, layanan, atau topik tertentu (Ulgasesa et al., 2022). Klasifikasi sentimen adalah proses untuk mengidentifikasi dan mengategorikan opini atau perasaan dalam teks, di mana model pembelajaran mesin digunakan untuk mengklasifikasikan sentimen tersebut, dengan metrik evaluasi seperti akurasi, presisi, recall, dan f1-score, sementara data teks yang digunakan biasanya melalui proses vectorization, seperti TF-IDF, untuk mengubah teks menjadi format yang dapat diproses oleh model (Limbong et al., 2022).

Klasifikasi sentimen adalah metode untuk mengukur perasaan atau pandangan seseorang terhadap suatu topik berdasarkan teks yang dihasilkan, seperti komentar, ulasan, atau opini, metode ini membantu mengidentifikasi apakah sentimen bersifat positif, negatif, atau netral, sehingga memudahkan pemahaman persepsi masyarakat terhadap topik atau layanan tertentu (Khairunniswa, 2026).

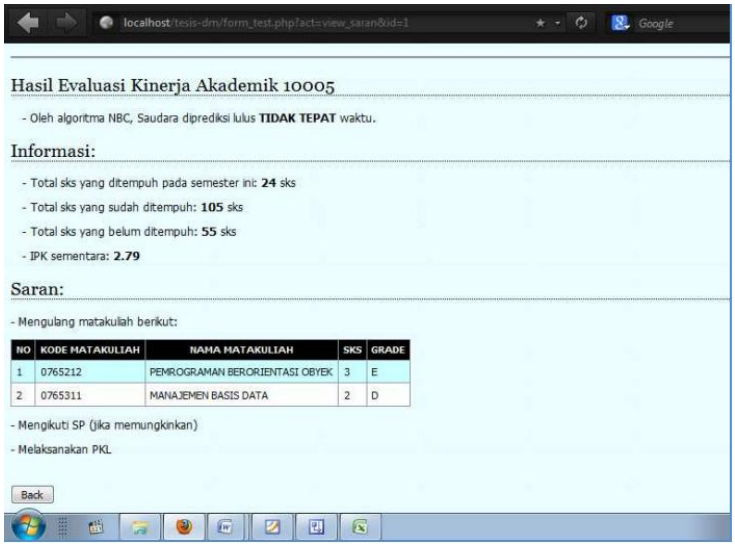
Dapat disimpulkan bahwa klasifikasi sentimen adalah teknik untuk mengidentifikasi dan mengelompokkan opini atau perasaan dalam teks, seperti komentar, ulasan, atau opini masyarakat. Proses ini menentukan apakah sentimen bersifat positif, negatif, atau netral, sehingga memudahkan pemahaman persepsi dan sikap penulis terhadap suatu topik, produk, atau layanan. Metode ini umumnya menggunakan model

pembelajaran mesin dan data teks yang telah melalui tahap preprocessing serta vectorization.

7. Evaluasi model Naive Bayes

Evaluasi model Naive Bayes adalah proses untuk menilai kemampuan model dalam mengklasifikasikan data baru setelah pelatihan. Proses ini menggunakan metrik seperti akurasi, presisi, recall, dan F1-score untuk mengukur kinerja model, memastikan bahwa prediksi terhadap data yang belum pernah dilihat sebelumnya tetap akurat dan andal (Sasmito et al., 2024). Evaluasi Model Naive Bayes merupakan tahap penting untuk menguji keakuratan model dalam memprediksi kategori data, di mana dengan menggunakan data uji yang terpisah dari data pelatihan, evaluasi ini mengevaluasi performa model berdasarkan metrik seperti akurasi dan f1-score, sehingga memastikan bahwa model Naive Bayes efektif dan dapat diandalkan dalam klasifikasi data nyata (Arisandi et al., 2024).

Evaluasi Model Naive Bayes adalah tahap untuk mengukur sejauh mana model Naive Bayes mampu memprediksi kategori data dengan benar, dengan membandingkan hasil prediksi pada data uji terhadap label yang sebenarnya, menggunakan metrik evaluasi seperti akurasi, presisi, recall, dan f1-score, yang membantu menilai performa model dalam mengklasifikasikan data secara tepat serta memastikan model tidak hanya menghafal data pelatihan tetapi juga generalizable ke data baru (Pajila & Sheena, 2023).



Gambar 2. 2 Evaluasi Model Naïve Bayes

Sumber : (Siskomti et. al., 2023)

Dapat disimpulkan bahwa evaluasi model Naive Bayes merupakan tahap penting untuk menilai kemampuan model dalam mengklasifikasikan data baru setelah pelatihan. Proses ini menggunakan metrik seperti akurasi, presisi, recall, dan F1-score untuk mengukur kinerja model, memastikan prediksi terhadap data yang belum pernah dilihat tetap akurat dan andal, serta menilai sejauh mana model dapat diterapkan secara efektif pada data nyata.

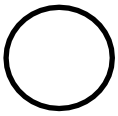
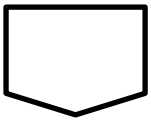
8. *Flowchart*

Flowchart merupakan diagram yang digunakan untuk menggambarkan alur proses atau langkah-langkah suatu sistem secara grafis menggunakan simbol-simbol tertentu. Flowchart membantu dalam memahami proses kerja suatu sistem secara lebih jelas dan sistematis.

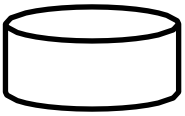
Dalam pengembangan perangkat lunak, flowchart digunakan untuk menggambarkan alur logika program atau proses yang terjadi dalam

sistem. Proses perancangan sistem dapat dilakukan secara lebih terstruktur sehingga memudahkan dalam tahap implementasi sistem, seperti pada tabel 2.1 Simbol–simbol flowchat.

Tabel 2.1. Simbol - simbol Flowchat

SIMBOL	NAMA	KETERANGAN
	*Input/Output*	Merepresentasikan *input* data atau *output* data yang diproses atau informasi
	Proses	Merepresentasikam operasi
	*Connector*	Keluar kea tau masuk dari bagian lain *flowchart* khususnya halaman yang sama
	*Off-page Connector*	Keluar kea tau masuk dari bagian lain *flowchart* halaman yang berbeda
	Anak Panah	Merepresentasikan alurkerja

SIMBOL	NAMA	KETERANGAN
	Keputusan	Keputusan dalam program
	*Preparation*	Pemberian harga awal
	*Terminal Point*	Awal atau akhir *flowchart*

	*Punched Card*	*Input atau Output* menggunakan kartu berlubang
	Dokumen	I/O dalam format yang dicetak
	Manual *Input*	*Input* yang dimasukkan manual dari *keyboard*
	*Database*	Menyimpan ke *database*

Sumber: (Hendrawansyah et al., 2026:55)

9. UML (*Unified Modeling Language*)

Unified Modeling Language (UML) merupakan bahasa pemodelan sistem perangkat lunak yang digunakan untuk memvisualisasikan, merancang, serta mendokumentasikan sistem agar proses pengembangan perangkat lunak dapat dilakukan secara lebih terstruktur (Abdiansyah et al., 2025:925). Menurut Pangestu & Voutama (2024:11846), UML memiliki keunggulan karena mampu memodelkan berbagai aspek dalam sistem perangkat lunak, mulai dari struktur, perilaku, hingga interaksi sistem. Selain itu, UML juga menjadi standar pemodelan yang digunakan secara luas sehingga dapat mempermudah komunikasi antara pengembang, perancang sistem, dan pihak terkait lainnya.

Tujuan penggunaan UML adalah menyediakan standar pemodelan dan istilah berorientasi objek yang dapat digunakan dalam menggambarkan proses pengembangan sistem mulai dari tahap analisis hingga implementasi

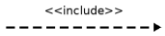
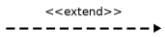
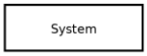
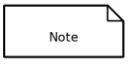
(Santoso & Migunani, 2021:37). Berikut ini diagram yang digunakan dalam perancangan sistem berorientasi objek.

a. Use Case Diagram

Dalam penelitian ini, penulis menggunakan Use Case Diagram sebagai pemodelan terhadap kelakuan (behavior) sistem yang akan dibuat. Use Case Diagram merupakan diagram yang digunakan untuk menggambarkan kebutuhan serta fungsi yang dapat dijalankan oleh suatu sistem (Destriana et al., 2021:7). Simbol-simbol Use Case Diagram dapat dilihat pada Tabel 2.2.

Tabel 2.2. Simbol-simbol Use Case Diagram

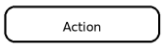
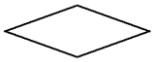
NO	GAMBAR	NAMA	KETERANGAN
1		Actor	Menjelaskan peran pengguna saat berinteraksi dengan use case dalam sistem.
2		Dependency	Hubungan ketika perubahan pada suatu elemen mandiri (independent) memengaruhi elemen lain yang bergantung padanya.
3		Generalization	Hubungan ketika objek turunan mewarisi struktur data dan perilaku dari objek induk.

NO	GAMBAR	NAMA	KETERANGAN
4		Include	Menunjukkan bahwa suatu use case secara eksplisit menggunakan use case lainnya.
5		Extend	Menunjukkan bahwa suatu use case dapat menambahkan atau memperluas perilaku use case lain pada kondisi tertentu.
6		Association	Menghubungkan satu objek dengan objek lainnya dalam sistem.
7		System	Menunjukkan paket atau bagian sistem dalam ruang lingkup tertentu.
8		Use Case	Menjelaskan urutan aktivitas sistem yang menghasilkan output bagi aktor.
9		Collaboration	Menunjukkan interaksi antar elemen dan aturan dalam sistem yang bekerja bersama untuk membentuk fungsi tertentu.
10		Note	Digunakan untuk memberikan catatan atau keterangan tambahan pada elemen diagram.

b. Activity Diagram

Activity Diagram merupakan diagram yang digunakan untuk memvisualisasikan alur aktivitas dalam sistem, mulai dari proses awal, keputusan yang terjadi selama proses berjalan, hingga aktivitas berakhir (Hasanah & Untari, 2020:79). Simbol-simbol Activity Diagram dapat dilihat pada Tabel 2.3.

Tabel 2.3. Simbol-simbol Activity Diagram

NO	GAMBAR	NAMA	KETERANGAN
1		Action	Keadaan sistem yang menggambarkan proses jalannya suatu aktivitas.
2		Initial Node	Menunjukkan titik awal atau proses inisialisasi aktivitas dalam sistem.
3		Activity Final Node	Menunjukkan titik akhir dari seluruh alur aktivitas dalam sistem.
4		Decision	Digunakan untuk menggambarkan keputusan atau tindakan berdasarkan kondisi tertentu.
5		Connector Line	Digunakan sebagai penghubung antar simbol dalam diagram.

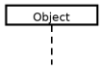
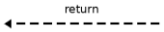
Sumber: (Yuliawati et al., 2024:284)

c. *Sequence Diagram*

Sequence Diagram merupakan salah satu diagram dalam UML yang digunakan untuk menggambarkan urutan proses dan interaksi antar objek dalam sistem berdasarkan alur waktu pemrosesan (Nabila et al.,

2021:82). Sequence Diagram memiliki beberapa simbol sebagaimana ditampilkan pada Tabel 2.4.

Tabel 2.4. Simbol-simbol Sequence Diagram

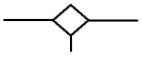
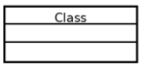
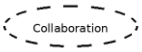
NO	GAMBAR	NAMA	KETERANGAN
1		LifeLine	Menunjukkan objek atau antarmuka yang saling berinteraksi dalam sistem selama proses berjalan.
2		Message	Menunjukkan proses komunikasi antar objek yang berisi informasi mengenai aktivitas atau proses yang terjadi dalam sistem.
3		Return Message	Menunjukkan pesan balasan atau komunikasi balik antar objek setelah suatu proses dijalankan.
4		Actor	Menunjukkan pihak atau entitas yang berinteraksi langsung dengan sistem aplikasi komputer.
5		<<Create>> Message	Menunjukkan pembuatan atau inisialisasi objek baru dalam sistem, dengan arah panah menuju objek yang dibuat.

Sumber: (Hardiyanti et al., 2021:162)

d. *Class Diagram*

Class Diagram merupakan diagram statis yang digunakan untuk menggambarkan struktur kelas beserta hubungan antar kelas dalam suatu sistem, termasuk atribut dan perilaku yang dimiliki oleh setiap kelas (Santoso & Migunani, 2021:186). Simbol-simbol Class Diagram dapat dilihat pada Tabel 2.5.

Tabel 2.5. Simbol-simbol Class Diagram

NO	GAMBAR	NAMA	KETERANGAN
1		Generalization	Hubungan pewarisan ketika objek turunan memiliki perilaku dan struktur data dari objek induk.
2		N-ary Association	Hubungan asosiasi yang melibatkan lebih dari dua objek dalam satu relasi.
3		Class	Kumpulan objek yang memiliki atribut dan operasi yang sama.
4		Collaboration	Interaksi antar elemen dalam sistem untuk menjalankan proses tertentu sehingga menghasilkan output bagi aktor.
5		Realization	Bentuk implementasi nyata dari operasi atau fungsi yang dijalankan oleh objek.
6		Dependency	Hubungan ketergantungan ketika perubahan pada satu elemen dapat memengaruhi elemen lain yang bergantung padanya.
7		Association	Hubungan yang menghubungkan satu objek dengan objek lainnya dalam sistem.

Sumber: (Yuliawati et al, 2024:284)

B. Kajian Empiris

Penelitian oleh Ramadhani et al. (2021), menunjukkan bahwa penerapan metode Naive Bayes efektif digunakan untuk analisis sentimen komentar masyarakat di media sosial. Hasil penelitian pada analisis komentar pelayanan publik menunjukkan akurasi model mencapai 84,5%. Penggunaan

preprocessing teks, tokenisasi, dan penghapusan stopword terbukti meningkatkan performa klasifikasi, sehingga metode ini sesuai untuk menilai opini masyarakat secara otomatis.

Penelitian oleh Prasetyo et al. (2022), menunjukkan bahwa penerapan metode Random Forest dan Naive Bayes dapat digunakan untuk mengklasifikasikan aduan pelanggan pada layanan publik. Hasil penelitian pada data aduan layanan kesehatan di Kabupaten Sleman menunjukkan Naive Bayes menghasilkan akurasi sebesar 87,3%, sedangkan Random Forest 89,1%. Hal ini menegaskan bahwa algoritma probabilistik efektif untuk sentimen analisis aduan publik.

Penelitian oleh Fadilah et al. (2023), menunjukkan bahwa penerapan preprocessing teks dan Naive Bayes dapat digunakan untuk mengklasifikasikan ulasan layanan publik digital. Penelitian pada dataset 2.000 komentar layanan publik di Kota Bandung menunjukkan bahwa model mencapai akurasi 86,7%. Hasil penelitian menegaskan bahwa langkah preprocessing teks sangat penting untuk meningkatkan performa model dalam analisis sentimen.

Penelitian oleh Nugroho et al. (2023), menunjukkan bahwa Naive Bayes Multinomial dapat digunakan untuk memprediksi opini pelanggan terhadap layanan publik atau produk. Penelitian pada ulasan online restoran menunjukkan nilai F1-score sebesar 0,82. Sistem ini terbukti mampu membedakan sentimen positif dan negatif secara efektif sehingga memberikan informasi yang berguna untuk perbaikan layanan.

Penelitian oleh Hidayat et al. (2024), menunjukkan bahwa penerapan Naive Bayes efektif untuk mengklasifikasikan data aduan publik berbasis web. Penelitian pada aduan layanan publik di Kota Surabaya menunjukkan akurasi model mencapai 88%. Sistem ini membantu instansi memahami keluhan masyarakat dengan lebih cepat, sekaligus mempermudah pengambilan keputusan terkait perbaikan layanan.

Penelitian oleh Putri dan Cahyono (2024) menunjukkan bahwa penerapan metode Multinomial Naive Bayes dapat digunakan untuk mengklasifikasikan komentar masyarakat terhadap pelayanan publik Pemerintah DKI Jakarta di media sosial Instagram. Penelitian ini membandingkan performa Support Vector Machine dan Multinomial Naive Bayes dalam mengklasifikasikan komentar masyarakat menjadi sentimen positif dan negatif, dengan tahap pemodelan yang mencakup pembagian data latih dan uji dalam berbagai rasio. Hasil penelitian menunjukkan SVM memperoleh akurasi 82%, sedangkan Multinomial Naive Bayes mencapai akurasi 80% dengan parameter alpha=2.0 dan fit_prior=True. Hal ini menegaskan bahwa Naive Bayes tetap kompetitif sebagai metode klasifikasi sentimen opini masyarakat terhadap layanan pemerintah.

Penelitian oleh Pratmanto, Imaniawan, dan Maarif (2023) menunjukkan bahwa metode Naive Bayes dan K-Nearest Neighbor dapat digunakan untuk menganalisis sentimen ulasan pengguna terhadap aplikasi layanan publik digital. Data ulasan aplikasi Identitas Kependudukan Digital (IKD) diambil dari Google Play Store, kemudian diproses melalui text cleaning,

case folding, tokenisasi, filtering, stemming, dan penghapusan stopword, sebelum ditransformasikan menjadi representasi numerik melalui pembobotan TF-IDF. Hasil penelitian menunjukkan bahwa KNN secara signifikan mengungguli Naive Bayes, dengan KNN mencapai akurasi rata-rata 82,85% dan presisi di atas 80% untuk kedua kelas sentimen. Penelitian ini menegaskan pentingnya tahap preprocessing teks dalam meningkatkan performa klasifikasi sentimen layanan publik digital.

Penelitian oleh Permadi (2020) menunjukkan bahwa metode Naive Bayes Classifier dapat digunakan untuk mengklasifikasikan sentimen ulasan pelanggan terhadap kualitas layanan. Penelitian ini mengklasifikasikan komentar pengunjung restoran di Singapura ke dalam dua kategori sentimen, yaitu positif (satisfied) dan negatif (unsatisfied), dengan hasil klasifikasi kemudian divisualisasikan melalui aplikasi berbasis web R Shiny berupa wordcloud dan diagram. Hasil pengujian menunjukkan algoritma Naive Bayes memberikan nilai akurasi sebesar 73%. Penelitian ini mengindikasikan bahwa Naive Bayes cukup efektif digunakan untuk menilai kepuasan pelanggan berdasarkan opini tekstual, meski akurasinya relatif moderat dibanding metode lain.

Penelitian oleh Nugroho, Hayati, dan Jabir (2025) menunjukkan bahwa metode Naive Bayes dan Random Forest dapat digunakan untuk membandingkan performa klasifikasi sentimen publik terhadap layanan digital pemerintah. Penelitian pada ulasan aplikasi Identitas Kependudukan Digital yang diambil dari Google Play Store ini melalui tahap pre-processing, pelabelan

berbasis lexicon, dan pembobotan TF-IDF. Hasil penelitian menunjukkan Random Forest memiliki performa lebih unggul dengan akurasi tertinggi 93%, sementara Naive Bayes mencapai akurasi 89%, dengan Random Forest juga menunjukkan performa yang lebih stabil pada precision dan recall untuk kedua kelas sentimen. Hal ini menegaskan bahwa meski Random Forest lebih unggul, Naive Bayes tetap menghasilkan akurasi yang tinggi untuk klasifikasi sentimen layanan publik digital.

Penelitian oleh Komarudin dan Hilda (2024) menunjukkan bahwa metode Naive Bayes efektif digunakan untuk menganalisis sentimen ulasan pengguna aplikasi layanan publik digital pada Play Store. Hasil penelitian menunjukkan model mencapai akurasi sebesar 82,23%, precision sebesar 76,08%, dan recall sebesar 94,02% berdasarkan perhitungan confusion matrix. Penelitian ini memperkuat temuan bahwa Naive Bayes mampu mengenali pola sentimen masyarakat terhadap layanan digital pemerintah dengan tingkat akurasi yang baik, sekaligus menjadi acuan praktis dalam pemilihan algoritma untuk analisis opini publik.

C. Kerangka Berfikir

Penelitian ini berfokus pada penerapan analisis sentimen terhadap data aduan pelayanan publik yang diterima oleh Dinas Kependudukan dan Pencatatan Sipil (Dukcapil) Kota Ngawi. Masalah utama yang ingin diselesaikan adalah bagaimana menganalisis dan mengklasifikasikan sentimen masyarakat yang terkandung dalam aduan terkait pelayanan yang diberikan oleh Dukcapil. Tujuan dari penelitian ini adalah untuk memberikan gambaran

yang lebih jelas mengenai tingkat kepuasan masyarakat melalui analisis sentimen dan untuk memberikan rekomendasi bagi perbaikan kualitas pelayanan berdasarkan temuan tersebut.

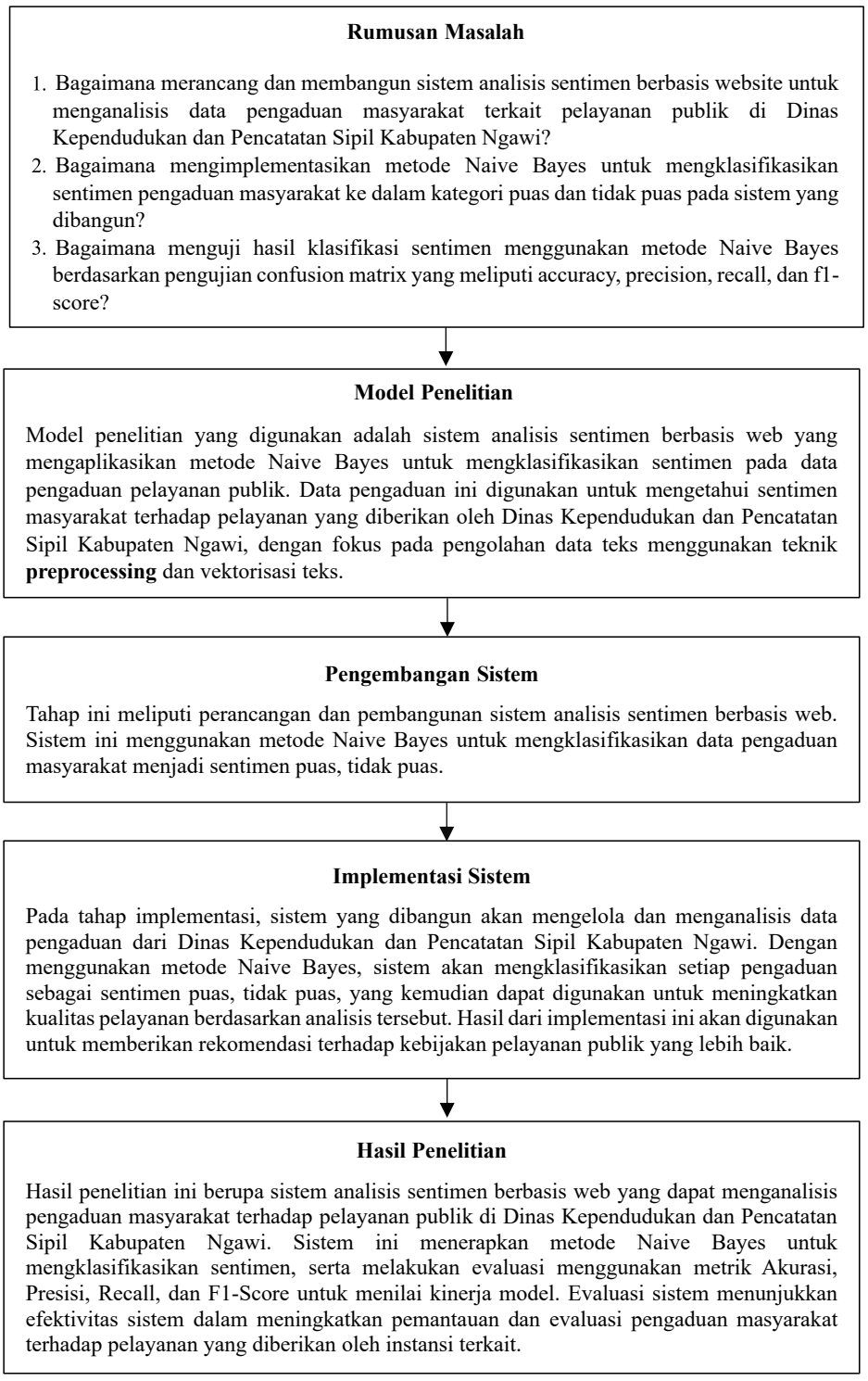
Fokus utama penelitian ini adalah bagaimana cara mengklasifikasikan sentimen dalam aduan pelayanan publik dan bagaimana hasil analisis tersebut dapat digunakan untuk meningkatkan proses pengambilan keputusan dalam pelayanan publik. Rumusan masalah yang diangkat dalam penelitian ini adalah: bagaimana cara mengklasifikasikan sentimen dalam data aduan pelayanan publik menggunakan metode Naive Bayes? dan bagaimana hasil analisis sentimen dapat memberikan rekomendasi dalam perbaikan kualitas layanan di Dukcapil Kota Ngawi.

Data yang akan digunakan dalam penelitian ini adalah aduan masyarakat yang diperoleh dari saluran pengaduan resmi yang disediakan oleh Dukcapil Kota Ngawi. Sebelum dianalisis lebih lanjut, data tersebut akan melalui tahap preprocessing, yang mencakup penghapusan karakter khusus, penghilangan stopwords, tokenisasi, dan stemming. Proses preprocessing ini bertujuan untuk membuat data teks lebih terstruktur dan siap digunakan dalam model klasifikasi.

Setelah tahap preprocessing, data teks akan diproses dengan menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency). TF-IDF akan mengubah teks menjadi representasi numerik yang bisa digunakan oleh model klasifikasi. Dengan menggunakan metode ini, kita dapat menilai kata-kata yang paling relevan dalam teks berdasarkan

frekuensinya dalam dokumen yang ada dan dalam keseluruhan dataset, sehingga mempermudah proses identifikasi sentimen yang terkandung dalam teks.

Model Naive Bayes akan digunakan untuk mengklasifikasikan sentimen dari data aduan. Model ini bekerja dengan menghitung probabilitas setiap kelas sentimen berdasarkan data pelatihan yang sudah ada. Sentimen akan dikategorikan menjadi tiga kelas: positif, negatif, dan netral. Proses ini akan membantu mengidentifikasi aduan yang perlu perhatian lebih atau yang menunjukkan kepuasan terhadap pelayanan yang diberikan.



Gambar 2. 3 Kerangka Berfikir