

Analisis Sentimen Berbasis Aspek Kinerja Pemerintah Kota Surabaya Menggunakan Algoritma Naïve Bayes

Ahmad Mustofa Akbar¹, Anita Qoiriah²

^{1,2} Program Studi S1 Teknik Informatika, Universitas Negeri Surabaya

¹ahmad.20107@mhs.unesa.ac.id

²anitagoiriah@unesa.ac.id

Abstrak—Dalam era digital, media sosial menjadi sarana yang sangat penting bagi masyarakat untuk menyuarakan opini terhadap kinerja pemerintah. Penelitian ini bertujuan menganalisis sentimen masyarakat terhadap kinerja Pemerintah Kota Surabaya berdasarkan aspek infrastruktur, ekonomi, pendidikan, dan kesehatan. Data dikumpulkan dari komentar publik di Instagram menggunakan teknik crawling dan dilabeli secara manual berdasarkan aspek dan polaritas sentimen. Setelah melalui tahapan pra-pemrosesan dan transformasi dengan TF-IDF, model dilatih menggunakan algoritma Multinomial Naïve Bayes. Hasil evaluasi menunjukkan performa model yang bervariasi pada setiap aspek. Pada aspek ekonomi, model mencapai akurasi 90%, dengan f1-score 0,93 (negatif) dan 0,82 (positif). Pada aspek kesehatan, akurasi sebesar 92%, dengan f1-score 0,93 (negatif) dan 0,90 (positif). Aspek pendidikan memperoleh akurasi 89%, dengan f1-score 0,92 (positif). Sementara itu, aspek infrastruktur memiliki performa terendah dengan akurasi 69%, dan f1-score hanya 0,45 untuk sentimen positif. Penelitian ini membuktikan bahwa pendekatan aspect-based sentiment analysis menggunakan Naïve Bayes efektif untuk mengidentifikasi persepsi masyarakat, serta memberikan kontribusi dalam pengambilan kebijakan berbasis data yang lebih responsif.

Kata Kunci— Analisis Sentimen, Aspect-Based Sentiment Analysis, Multinomial Naïve Bayes, Media Sosial, Infrastruktur, Ekonomi, Pendidikan, Kesehatan, Pemerintah Kota Surabaya.

I. PENDAHULUAN

Surabaya merupakan salah satu kota metropolitan di Indonesia yang memiliki peran penting dalam dinamika pemerintahan daerah, khususnya dalam hal penyediaan layanan publik yang menyentuh berbagai sektor strategis seperti infrastruktur, pendidikan, kesehatan, dan ekonomi. Dalam konteks perkembangan era digital, media sosial telah menjadi sarana interaktif yang memungkinkan masyarakat menyampaikan aspirasi, kritik, maupun dukungan terhadap kebijakan pemerintah secara terbuka dan real-time. Pemanfaatan media sosial sebagai kanal komunikasi publik juga diakui berkontribusi dalam membentuk persepsi warga terhadap prinsip *good governance* yang dijalankan oleh pemerintah daerah [1].

Sentimen publik yang terkandung dalam komentar-komentar di media sosial dapat mencerminkan berbagai bentuk persepsi dan evaluasi terhadap kinerja pemerintah. Oleh karena itu, penting bagi pemerintah daerah seperti Kota Surabaya untuk memahami opini masyarakat secara sistematis agar kebijakan yang disusun berbasis pada kebutuhan dan respons aktual dari masyarakat. Salah satu pendekatan yang relevan

dalam konteks ini adalah *Aspect-Based Sentiment Analysis* (ABSA) yaitu metode analisis yang tidak hanya mengidentifikasi polaritas opini (positif atau negatif), tetapi juga mengaitkannya dengan aspek spesifik seperti kualitas infrastruktur, pelayanan kesehatan, mutu pendidikan, dan kebijakan ekonomi [2][3].

Untuk mengolah data dalam jumlah besar secara efisien, diperlukan algoritma klasifikasi yang mampu bekerja cepat dan akurat. Algoritma *Naïve Bayes* merupakan salah satu metode klasifikasi berbasis probabilistik yang umum digunakan dalam analisis teks karena kemampuannya dalam memproses data skala besar dengan asumsi independensi antar fitur [4]. Keunggulan metode ini juga telah ditunjukkan dalam berbagai studi terdahulu. Toy et al. [5] menerapkan *Naïve Bayes* dalam klasifikasi opini publik terhadap kebijakan *new normal*, dan memperoleh akurasi mencapai 73,3% pada data uji. Sementara itu, Rahayu et al. [3] melaporkan bahwa algoritma ini mampu mengklasifikasikan sentimen terhadap program Kampus Merdeka dengan akurasi sebesar 86%, presisi 87%, dan recall 80%. Keberhasilan lain juga ditemukan dalam penelitian Rivanie et al. [6] yang menganalisis opini publik terhadap Menteri Kesehatan selama pandemi COVID-19, menunjukkan efektivitas metode ini dalam konteks kebijakan publik.

Penelitian ini bertujuan untuk menerapkan algoritma *Naïve Bayes* dalam menganalisis sentimen masyarakat terhadap kinerja Pemerintah Kota Surabaya berdasarkan empat aspek utama: infrastruktur, pendidikan, kesehatan, dan ekonomi. Dua rumusan masalah utama diangkat dalam penelitian ini, yaitu: (1) bagaimana penerapan algoritma *Naïve Bayes* untuk mengidentifikasi sentimen publik terhadap aspek-aspek kinerja Pemerintah Kota Surabaya, dan (2) bagaimana hasil akurasi dari model klasifikasi sentimen yang dikembangkan.

Tujuan dari penelitian ini adalah untuk mengembangkan model analisis sentimen berbasis aspek yang dapat digunakan untuk mengklasifikasikan opini masyarakat secara lebih terstruktur dan akurat, serta menghasilkan ukuran kinerja model dalam bentuk metrik evaluasi seperti akurasi dan f1-score. Penelitian ini diharapkan memberikan kontribusi praktis dalam menyediakan data berbasis opini publik yang dapat digunakan oleh pemerintah daerah dalam meningkatkan kualitas layanan dan menyusun kebijakan yang lebih responsif.

Manfaat utama dari penelitian ini meliputi: (1) memberikan gambaran persepsi masyarakat terhadap berbagai sektor kinerja pemerintah Kota Surabaya; (2) menjadi sumber data bagi pembuat kebijakan untuk menyusun langkah strategis yang lebih terarah dan berbasis opini publik; (3) memperkuat transparansi dan akuntabilitas pemerintah daerah melalui

keterlibatan warga secara digital; serta (4) menjadi referensi metodologis untuk penelitian serupa di kota atau daerah lain [3][7].

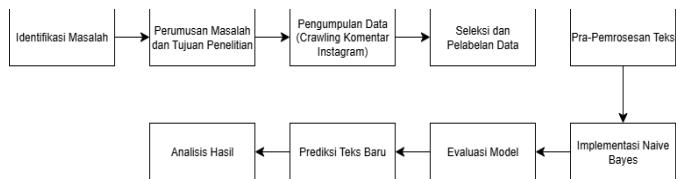
Lingkup penelitian ini dibatasi pada analisis sentimen terhadap empat aspek kinerja pemerintah: infrastruktur, pendidikan, kesehatan, dan ekonomi. Data yang dianalisis diperoleh dari komentar publik di platform Instagram, yang telah melalui proses kurasi dan pelabelan manual untuk memastikan relevansi dan validitasnya. Komentar bersifat netral atau tidak berhubungan dengan aspek yang diteliti tidak disertakan dalam analisis lebih lanjut [4].

Dengan menggunakan pendekatan ABSA dan algoritma *Naïve Bayes*, penelitian ini diharapkan mampu memberikan pemetaan sentimen publik yang tajam dan akurat, serta menjadi kontribusi nyata dalam pengembangan kebijakan berbasis data (*evidence-based policy*).

II. METODE PENELITIAN

A. Desain dan Alur Penelitian

Penelitian ini dirancang dengan pendekatan kuantitatif menggunakan metode klasifikasi teks berbasis algoritma Naïve Bayes. Fokus utama penelitian adalah membangun model *Aspect-Based Sentiment Analysis* (ABSA) terhadap opini masyarakat yang ditujukan pada kinerja Pemerintah Kota Surabaya melalui data komentar media sosial. Tahapan pelaksanaan penelitian dilakukan secara berurutan dimulai dari: (1) identifikasi dan perumusan masalah; (2) pengumpulan data dari media sosial Instagram; (3) seleksi dan pelabelan data berdasarkan aspek dan polaritas; (4) pra-pemrosesan data teks; (5) transformasi data ke dalam bentuk numerik menggunakan TF-IDF; (6) pelatihan model klasifikasi menggunakan algoritma *Naïve Bayes*; dan (7) evaluasi model dengan menggunakan metrik akurasi, precision, recall, dan f1-score. Alur lengkap dari proses penelitian divisualisasikan dalam bentuk diagram alir pada gbr 1 yang menjadi dasar struktur kerja penelitian ini.



Gbr 1. Diagram Alur Penelitian

B. Pengumpulan dan Seleksi Data

Data yang digunakan dalam penelitian ini diperoleh dari komentar publik pada unggahan di platform media sosial Instagram, dengan fokus utama pada akun-akun yang memiliki kaitan langsung dengan kinerja Pemerintah Kota Surabaya. Komentar dikumpulkan melalui proses *web crawling* menggunakan ekstensi *IG Comment Export*, yang memungkinkan ekstraksi komentar dalam format CSV. Target pengambilan data mencakup akun resmi Pemerintah Kota Surabaya, dinas-dinas layanan publik, serta media lokal yang

sering mengunggah konten terkait kebijakan dan program pemerintah.

Dari hasil crawling awal, diperoleh sekitar 3.500 komentar mentah. Namun, tidak semua data dapat langsung digunakan karena sebagian mengandung noise seperti spam, komentar kosong, promosi, atau komentar yang tidak relevan dengan aspek yang diteliti. Oleh karena itu, dilakukan proses seleksi data berdasarkan kriteria sebagai berikut:

- Komentar berbentuk teks (bukan emoji, gambar, atau simbol);
- Memuat opini terhadap kebijakan, layanan, atau program pemerintah;
- Memiliki keterkaitan eksplisit atau implisit dengan aspek infrastruktur, pendidikan, kesehatan, atau ekonomi;
- Bersifat publik dan legal untuk diakses.

Komentar yang bersifat netral informatif, promosi, atau hanya terdiri dari kata-kata seperti “oke” dan “mantap” juga dieliminasi. Hasil dari proses seleksi ini menghasilkan sebanyak 2.563 komentar yang layak digunakan sebagai data inti penelitian. Pendekatan seleksi ini mengacu pada praktik validasi manual yang lazim digunakan dalam studi analisis sentimen untuk menjaga kualitas data input ke dalam model klasifikasi [1][2].

C. Pelabelan Data

Pelabelan data dilakukan secara manual dengan meninjau isi komentar satu per satu. Proses ini dilakukan oleh peneliti secara langsung untuk mengidentifikasi makna dan konteks dari setiap komentar. Pelabelan dilakukan berdasarkan dua dimensi utama, yaitu:

1. Aspek: komentar dikategorikan ke dalam salah satu dari empat aspek utama kinerja pemerintah, yaitu infrastruktur, pendidikan, kesehatan, dan ekonomi. Penentuan aspek dilakukan berdasarkan kata kunci dan konteks pembahasan yang terdapat dalam komentar.
2. Polaritas Sentimen: setiap komentar diberi label sebagai sentimen positif atau negatif. Sentimen positif menunjukkan apresiasi, puji, atau dukungan terhadap pemerintah, sedangkan sentimen negatif mencerminkan kritik, keluhan, atau ketidakpuasan.

Awalnya, sentimen netral juga ikut dilabeli untuk mewakili komentar informatif atau tidak mengandung ekspresi emosional. Namun, setelah dilakukan analisis awal, ditemukan bahwa keberadaan kelas netral justru menurunkan performa klasifikasi dan memperburuk distribusi data antar kelas. Hal ini sejalan dengan studi sebelumnya yang menyarankan penghapusan kelas netral untuk menghindari noise dan ketidakseimbangan kelas [1][3]. Oleh karena itu, dalam tahap akhir, komentar dengan sentimen netral dihapus dan proses klasifikasi difokuskan hanya pada dua kelas utama: positif dan negatif.

Verifikasi ulang pelabelan dilakukan untuk memastikan konsistensi dan keakuratan label. Komentar yang ambigu, multitopik, atau tidak jelas konteksnya juga dieliminasi dari dataset akhir. Hasil akhir pelabelan menjadi fondasi penting

dalam pelatihan model, karena secara langsung memengaruhi validitas dan reliabilitas model klasifikasi yang dibangun.

D. Pra-pemrosesan Teks

Tahap pra-pemrosesan teks dilakukan untuk membersihkan dan menormalkan data komentar yang masih dalam bentuk mentah. Proses ini penting agar data teks siap digunakan dalam proses ekstraksi fitur dan pelatihan model klasifikasi. Langkah-langkah yang diterapkan pada penelitian ini meliputi beberapa tahapan standar dalam pengolahan bahasa alami (*Natural Language Processing/NLP*) berbahasa Indonesia, yaitu:

1. Case Folding

Semua huruf dalam komentar diubah menjadi huruf kecil untuk menyamakan representasi kata yang secara semantik sama tetapi secara teknis berbeda, seperti “Surabaya” dan “surabaya”.

2. Pembersihan Teks (Text Cleaning)

Seluruh karakter *non-teks* yang tidak relevan dihapus, meliputi:

- Tautan atau URL (contoh: <https://...>);
- Simbol dan karakter khusus (misalnya @, #, !, &, dll);
- Angka dan tanda baca;
- *Whitespace* berlebih.

Pembersihan dilakukan menggunakan ekspresi reguler (*regex*) dan fungsi pemrosesan string di Python.

3. Tokenisasi

Teks yang telah dibersihkan dipecah menjadi unit kata (token) menggunakan pustaka NLP Python. Tokenisasi membantu dalam identifikasi kata-kata penting yang digunakan dalam proses pembobotan dan analisis.

4. Stopword Removal

Kata-kata umum (seperti “yang”, “dan”, “di”) yang tidak memiliki makna signifikan dalam analisis sentimen dihapus menggunakan daftar *stopwords* Bahasa Indonesia dari pustaka *Sastrawi*.

5. Stemming

Proses stemming dilakukan untuk mengembalikan setiap kata ke bentuk dasarnya. Sebagai contoh, kata “mengajar”, “pengajaran”, dan “diajarkan” semuanya akan dikembalikan ke bentuk dasar “ajar”. Proses ini membantu mengurangi keragaman kata dan memperkuat representasi semantik.

Langkah-langkah pra-pemrosesan tersebut bertujuan untuk menghasilkan korpus data yang bersih, terstruktur, dan siap untuk dikonversi ke dalam bentuk numerik dalam tahap berikutnya [2][4].

E. Vektorisasi dan Pelatihan Model

Setelah teks diproses, data kemudian dikonversi ke dalam representasi numerik menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF digunakan untuk menghitung bobot pentingnya suatu kata dalam sebuah

dokumen relatif terhadap seluruh korpus. Kata-kata yang lebih sering muncul dalam satu komentar tetapi jarang muncul di komentar lain akan memiliki bobot lebih tinggi, sehingga dianggap lebih informatif.

Output dari TF-IDF adalah matriks nilai numerik yang menggambarkan seberapa penting setiap kata (fitur) terhadap setiap dokumen (komentar). Matriks ini menjadi input untuk pelatihan model klasifikasi sentimen.

Model dikembangkan menggunakan algoritma *Multinomial Naïve Bayes*, yang merupakan varian dari Naïve Bayes yang cocok untuk data dengan fitur diskrit seperti frekuensi kata. Untuk menguji performa model, dilakukan lima skenario pembagian data pelatihan dan pengujian (*train-test split*), yaitu: 50:50, 60:40, 70:30, 80:20, dan 90:10. Proses pelatihan model dilakukan menggunakan pustaka *Scikit-learn* di Python. Setiap skenario diuji secara independen untuk melihat pengaruh proporsi data terhadap akurasi dan stabilitas performa model [5].

F. Evaluasi Kinerja Model

Model yang telah dilatih dievaluasi menggunakan metrik standar dalam klasifikasi biner, yaitu:

- Akurasi: Proporsi prediksi yang benar dari seluruh jumlah data;
- Precision: Proporsi prediksi positif yang benar-benar relevan;
- Recall: Proporsi data positif yang berhasil dikenali dengan benar;
- F1-Score: Harmonic mean dari precision dan recall, digunakan untuk menyeimbangkan keduanya.

Evaluasi dilakukan melalui dua pendekatan utama:

1. Classification Report

Laporan ini menunjukkan nilai precision, recall, f1-score, dan support untuk masing-masing kelas sentimen (positif dan negatif).

2. Confusion Matrix

Matrix ini menggambarkan jumlah prediksi yang benar dan salah berdasarkan kategori aktual dan prediksi, memberikan gambaran menyeluruh tentang kesalahan klasifikasi.

Hasil evaluasi menunjukkan bahwa performa model bervariasi antar skenario, dengan f1-score tertinggi sebesar 0,90 diperoleh pada skenario 90:10 untuk aspek infrastruktur. Hal ini menunjukkan bahwa semakin banyak data pelatihan yang digunakan, semakin tinggi pula akurasi klasifikasi yang dapat dicapai. Secara umum, aspek infrastruktur dan ekonomi memberikan hasil klasifikasi yang lebih stabil dibandingkan aspek pendidikan dan kesehatan [1][6].

III. HASIL DAN PEMBAHASAN

A. Statistik Dataset dan Distribusi Aspek

Dataset yang digunakan dalam penelitian ini terdiri dari 2.663 komentar publik yang diperoleh dari media sosial Instagram, yang telah dikategorikan berdasarkan empat aspek

utama kinerja Pemerintah Kota Surabaya: ekonomi, infrastruktur, kesehatan, dan pendidikan.

TABEL I
DISTRIBUSI JUMLAH KOMENTAR BERDASARKAN ASPEK

No	Aspek	Jumlah Komentar	Percentase (%)
1	Ekonomi	1.077	40,45%
2	Infrastruktur	688	25,84%
3	Kesehatan	596	22,39%
4	Pendidikan	302	11,34%
	Total	2.663	100%

Distribusi komentar yang ditunjukkan pada Tabel 1 menunjukkan bahwa aspek ekonomi merupakan kategori yang paling banyak dibahas oleh pengguna media sosial, dengan total 1.077 komentar atau sekitar 40,45% dari keseluruhan data. Komentar-komentar pada aspek ini umumnya berkaitan dengan kondisi ekonomi daerah, bantuan sosial, biaya hidup, serta ketersediaan lapangan pekerjaan, mencerminkan kepekaan tinggi masyarakat terhadap dinamika ekonomi daerah.

Aspek kedua yang paling banyak muncul adalah infrastruktur dengan 688 komentar (25,84%), yang mencerminkan perhatian publik terhadap layanan fisik seperti jalan, transportasi, penerangan, dan fasilitas umum lainnya. Disusul oleh aspek kesehatan sebanyak 596 komentar (22,39%), yang mencerminkan kepedulian masyarakat terhadap layanan kesehatan, vaksinasi, ketersediaan fasilitas medis, serta respons pemerintah terhadap pandemi.

Sementara itu, aspek pendidikan tercatat sebagai yang paling sedikit mendapatkan perhatian, hanya sebanyak 302 komentar (11,34%). Hal ini dimungkinkan karena minimnya topik pendidikan yang menjadi viral atau terbatasnya kebijakan pendidikan lokal yang menjadi sorotan selama periode pengambilan data.

Distribusi ini memberikan gambaran awal bahwa isu ekonomi dan infrastruktur menjadi dominan dalam persepsi publik terhadap kinerja pemerintah, sedangkan isu pendidikan cenderung kurang mendapat porsi diskusi dalam ruang digital, meskipun memiliki urgensi sendiri.

Distribusi ini menjadi dasar penting dalam proses pelabelan dan pelatihan model klasifikasi, karena proporsi yang tidak seimbang antar aspek berpotensi memengaruhi performa algoritma, khususnya dalam hal presisi dan sensitivitas terhadap aspek minor seperti pendidikan [7].

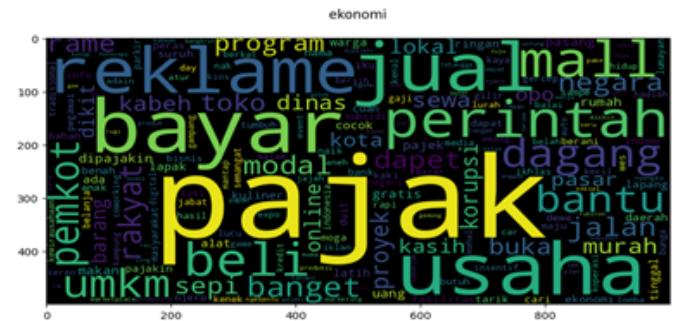
B. Visualisasi Wordcloud

Wordcloud digunakan untuk menggambarkan kata-kata yang paling sering muncul dalam komentar publik untuk masing-masing aspek. Semakin besar ukuran suatu kata dalam wordcloud, semakin sering kata tersebut muncul dalam komentar yang dikumpulkan.

1. Aspek Ekonomi

Pada Gbr 2, terlihat bahwa kata “*pajak*”, “*usaha*”, dan “*bayar*” mendominasi wordcloud untuk aspek ekonomi. Hal ini menunjukkan bahwa publik banyak memberikan perhatian terhadap isu-isu fiskal dan kewirausahaan. Kata-kata seperti “*reklame*”, “*mall*”, dan “*sewa*” juga mencerminkan diskusi

tentang aktivitas perdagangan dan biaya operasional yang menjadi beban bagi pelaku usaha.



Gbr 2. Wordcloud Komentar Aspek Ekonomi

2. Aspek Infrastruktur

Gbr 3 menunjukkan bahwa kata “*jalan*”, “*aspal*”, “*baik*”, dan “*surabaya*” paling sering digunakan pada aspek infrastruktur. Ini mencerminkan bahwa kualitas dan kondisi infrastruktur fisik seperti jalan raya menjadi perhatian utama. Kata-kata seperti “*gorong*”, “*macet*”, dan “*lampu*” juga menunjukkan bahwa masyarakat menyoroti fasilitas umum lainnya yang memengaruhi mobilitas dan kenyamanan warga kota.



Gbr 3. Wordcloud Komentar Aspek Infrastruktur

3. Aspek Kesehatan

Dalam Gbr 4, kata “*layanan*”, “*alat*”, dan “*canggih*” menjadi dominan. Ini mengindikasikan bahwa masyarakat memperhatikan kelengkapan fasilitas medis, kualitas alat kesehatan, serta kecanggihan layanan yang diberikan. Kata lain seperti “*dokter*”, “*bpjs*”, dan “*vaksin*” juga menunjukkan fokus pada sistem pelayanan kesehatan dan aksesibilitas program kesehatan.



Gbr 4. Wordcloud Komentar Aspek Kesehatan

	Akurasi		0.88	113
--	---------	--	------	-----

4. Aspek Pendidikan

Sementara itu, Gbr 5 memperlihatkan bahwa kata “anak”, “mbr” (Masyarakat Berpenghasilan Rendah), dan “sekolah” mendominasi pada aspek pendidikan. Hal ini menunjukkan bahwa publik paling sering membicarakan kebijakan pendidikan inklusif, bantuan seragam, serta akses pendidikan untuk kelompok rentan. Kata seperti “program”, “bantu”, dan “guru” juga muncul sebagai penanda bahwa masyarakat menyoroti inisiatif pemerintah dalam bidang pendidikan dasar.



Gbr 5. Wordcloud Komentar Aspek Pendidikan

C. Hasil Klasifikasi Sentimen dengan Naïve Bayes

Untuk mengetahui pengaruh proporsi data latih dan uji terhadap performa model klasifikasi, penelitian ini mengimplementasikan lima skenario pembagian dataset, yakni 50:50, 60:40, 70:30, 80:20, dan 90:10. Dataset yang digunakan merupakan gabungan dari semua aspek komentar (ekonomi, infrastruktur, kesehatan, dan pendidikan) yang telah melalui proses pelabelan sentimen (positif dan negatif) dan transformasi fitur menggunakan teknik TF-IDF [9].

Evaluasi dilakukan menggunakan metrik *precision*, *recall*, dan *f1-score* untuk masing-masing kelas sentimen, serta *overall accuracy* untuk setiap skenario. Ringkasan hasil pengujian disajikan dalam Tabel 2 berikut:

TABEL II
 HASIL EVALUASI MODEL NAÏVE BAYES BERDASARKAN SKENARIO PEMBAGIAN DATA

Skenario	Sentimen	Precision	Recall	F1-Score	Support
50:50	Negatif	0.79	0.96	0.86	325
	Positif	0.91	0.63	0.75	231
	Akurasi			0.82	556
60:40	Negatif	0.80	0.97	0.87	260
	Positif	0.93	0.66	0.77	186
	Akurasi			0.84	446
70:30	Negatif	0.81	0.97	0.88	196
	Positif	0.94	0.68	0.79	139
	Akurasi			0.85	335
80:20	Negatif	0.81	0.97	0.88	196
	Positif	0.94	0.68	0.79	139
	Akurasi			0.85	335
90:10	Negatif	0.87	0.94	0.91	71
	Positif	0.89	0.76	0.82	42
	Akurasi				

Hasil pengujian menunjukkan bahwa performa model meningkat secara konsisten seiring bertambahnya proporsi data latih. Pada skenario 90:10, model mencapai *f1-score* tertinggi sebesar 0.91 untuk sentimen negatif dan 0.82 untuk sentimen positif, dengan akurasi keseluruhan sebesar 88%.

Secara umum:

- *Recall* untuk sentimen negatif sangat tinggi di semua skenario (≥ 0.94), menandakan bahwa model sangat sensitif dalam mengenali komentar negatif.
- *Precision* untuk sentimen positif mencapai nilai tertinggi pada skenario 70:30 dan 80:20 (0.94), menunjukkan keakuratan tinggi saat model memprediksi komentar bernada positif.
- Nilai *f1-score* untuk kedua sentimen meningkat secara konsisten dari skenario 50:50 hingga 90:10, mencerminkan kestabilan dan ketepatan klasifikasi model.

Temuan ini memperkuat dugaan bahwa porsi data latih yang lebih besar menghasilkan model yang lebih akurat dan seimbang, terutama dalam konteks data opini publik yang cenderung tidak terdistribusi secara merata antar kelas [8].

D. Evaluasi Model: F1-Score dan Confusion Matrix

Evaluasi kinerja model klasifikasi tidak hanya dilakukan secara agregat berdasarkan skenario pembagian data, tetapi juga secara mendetail berdasarkan aspek kinerja pemerintah: infrastruktur, ekonomi, pendidikan, dan kesehatan. Penilaian ini menggunakan metrik evaluasi umum, yaitu *precision*, *recall*, dan *f1-score*, yang diperoleh melalui *classification report*, serta matriks kebingungan (*confusion matrix*) untuk mengidentifikasi jenis kesalahan klasifikasi yang terjadi [10].

1. Classification Report per Aspek

Hasil *classification report* pada tabel 3 dibawah menunjukkan bahwa aspek ekonomi dan kesehatan memperoleh performa model yang paling tinggi. Pada aspek ekonomi, model mencapai *f1-score* 0.93 untuk sentimen negatif dan 0.82 untuk sentimen positif, dengan akurasi keseluruhan sebesar 90%. Hal serupa juga terjadi pada aspek kesehatan, di mana *f1-score* untuk sentimen negatif mencapai 0.93, dan untuk sentimen positif sebesar 0.90, dengan akurasi total sebesar 92%.

Aspek pendidikan juga menunjukkan performa yang baik, dengan *f1-score* 0.92 pada sentimen positif, meskipun jumlah data uji pada aspek ini relatif kecil (total 35 data). Sebaliknya, aspek infrastruktur memiliki performa paling rendah, dengan akurasi keseluruhan hanya 69%, terutama disebabkan oleh rendahnya recall untuk sentimen positif (0.31) yang menunjukkan banyak komentar positif tidak dikenali dengan baik oleh model.

TABEL III
 CLASSIFICATION REPORT PER ASPEK

Aspek	Sentimen	Precision	Recall	F1-Score	Support
Infrastruktur	Negatif	0.66	0.96	0.78	51
	Positif	0.85	0.31	0.45	36

	Akurasi			0.69	87
Ekonomi	Negatif	0.87	1.00	0.93	95
	Positif	1.00	0.69	0.82	45
	Akurasi			0.90	140
Pendidikan	Negatif	1.00	0.64	0.78	11
	Positif	0.86	1.00	0.92	24
	Akurasi			0.89	35
Kesehatan	Negatif	0.87	1.00	0.93	39
	Positif	1.00	0.82	0.90	34
	Akurasi			0.92	73
Rata-rata	Negatif	0.81	0.97	0.88	196
	Positif	0.94	0.68	0.79	139
	Akurasi			0.85	335

2. Confusion Matrix per Aspek

Untuk melihat jenis kesalahan klasifikasi secara eksplisit, digunakan *confusion matrix* yang memperlihatkan jumlah prediksi benar dan salah dari masing-masing kelas. Matriks ini memberikan informasi terkait *False Positive* (FP) dan *False Negative* (FN), yang penting dalam konteks analisis opini publik [10].

Sebagai contoh, pada aspek infrastruktur, terdapat 25 komentar positif yang diklasifikasikan sebagai negatif (FN), sehingga recall untuk kelas positif menjadi sangat rendah. Sebaliknya, aspek ekonomi dan kesehatan menunjukkan akurasi prediksi yang tinggi dengan jumlah kesalahan minimal. Pada aspek pendidikan, performa klasifikasi terhadap komentar positif sangat baik, dengan tidak ada *False Negative* (FN).

TABEL IV
 CONFUSION MATRIX PER ASPEK

Aspek	TN	FP	FN	TP
Infrastruktur	49	2	25	11
Ekonomi	95	0	14	31
Pendidikan	7	4	0	24
Kesehatan	39	0	6	28
Total	190	6	95	94

Interpretasi dari tabel 4 *confusion matrix* gabungan (baris terakhir) memperlihatkan bahwa model menghasilkan *True Negative* (TN) yang tinggi, tetapi juga cukup banyak *False Negative* (FN), yaitu komentar positif yang gagal dikenali sebagai positif. Hal ini sejalan dengan temuan pada *classification report*, di mana *recall* untuk kelas positif relatif lebih rendah dibandingkan kelas negatif [8].

E. Analisis dan Pembahasan

Hasil klasifikasi sentimen terhadap komentar publik mengenai kinerja Pemerintah Kota Surabaya menunjukkan bahwa model *Multinomial Naïve Bayes* berbasis TF-IDF berhasil mengidentifikasi opini masyarakat dengan cukup akurat. Pada skenario pembagian data 90:10, model mencatatkan akurasi tertinggi sebesar 88%, dengan *f1-score* masing-masing 0.91 untuk kelas negatif dan 0.82 untuk kelas positif. Temuan ini menunjukkan bahwa proporsi data latih yang besar secara langsung meningkatkan generalisasi model, sebagaimana diperkuat dalam penelitian terdahulu yang menunjukkan hubungan antara jumlah data pelatihan dan akurasi klasifikasi sentimen [1].

Performa terbaik model ditunjukkan pada aspek ekonomi dan kesehatan. Pada aspek ekonomi, model menghasilkan *f1-score* 0.93 (negatif) dan 0.82 (positif), dengan akurasi keseluruhan sebesar 90%. Demikian pula, aspek kesehatan mencatat *f1-score* 0.93 (negatif) dan 0.90 (positif), serta akurasi total 92%. Keberhasilan ini didukung oleh pola kosakata yang konsisten, seperti istilah bantuan sosial, biaya hidup, vaksinasi, dan BPJS, yang mempermudah algoritma dalam membedakan polaritas komentar [5].

Sebaliknya, aspek infrastruktur menunjukkan performa klasifikasi terendah. Meskipun memiliki jumlah data yang besar, akurasi klasifikasi hanya mencapai 69%, dengan *recall* untuk kelas positif sangat rendah (0.31). Hal ini menyebabkan *f1-score* positif hanya sebesar 0.45. Rendahnya performa ini berkaitan dengan kosakata komentar yang netral atau ambigu, seperti “lumayan baik” atau “sudah bagus”, yang tidak memberikan sinyal sentimen yang kuat bagi model berbasis probabilistik [3].

Pada aspek pendidikan, model menunjukkan performa yang relatif baik meskipun jumlah datanya terbatas. Akurasi klasifikasi mencapai 89%, dengan *f1-score* 0.92 untuk komentar positif. Keberhasilan ini dikaitkan dengan kejelasan dan spesifiknya kosakata dalam komentar, misalnya “kurikulum”, “PPDB”, atau “beasiswa”, yang memperkuat kemampuan klasifikasi model [6].

Secara keseluruhan, hasil ini mengonfirmasi bahwa algoritma sederhana seperti *Naïve Bayes* tetap dapat digunakan secara efektif dalam klasifikasi opini publik berbasis aspek, asalkan didukung oleh pra-pemrosesan teks yang menyeluruh dan transformasi fitur yang representatif seperti TF-IDF [4]. Dalam konteks ini, kualitas data dan konsistensi pola ujaran menjadi faktor krusial untuk meningkatkan performa klasifikasi.

Secara agregat, model menunjukkan kecenderungan untuk mendekripsi sentimen negatif dengan lebih baik dibanding sentimen positif. Hal ini terlihat dari nilai *recall* dan *f1-score* yang konsisten lebih tinggi untuk kelas negatif pada hampir semua aspek. Sebagai contoh, pada aspek infrastruktur, recall untuk komentar negatif mencapai 0.96, sementara untuk komentar positif hanya 0.31. Ketidakseimbangan ini mengindikasikan adanya bias dalam model terhadap komentar yang mengandung kata-kata keluhan, kritik, atau ekspresi negatif yang cenderung eksplisit [5].

Keberhasilan klasifikasi pada aspek ekonomi dan kesehatan dapat pula dijelaskan melalui kualitas dan kuantitas data pelatihan yang tersedia. Sebagaimana dijelaskan oleh Rahayu et al. [3], akurasi model sangat dipengaruhi oleh representasi fitur yang kuat dan keberagaman data latih. Kosakata pada komentar ekonomi dan kesehatan relatif spesifik dan berulang (misalnya: “harga naik”, “bantuan telat”, “rumah sakit penuh”), sehingga model lebih mudah mengenali pola distribusi kata tersebut. Sebaliknya, komentar positif sering kali menggunakan kata-kata yang bersifat netral atau mengandung puji-pujian tidak langsung, yang menyulitkan model dalam mengklasifikasikannya.

Kinerja model juga menunjukkan bahwa ukuran dataset bukan satu-satunya faktor penentu performa. Meskipun aspek pendidikan memiliki jumlah data terkecil, model justru berhasil mengklasifikasikan komentar dengan akurasi tinggi. Hal ini sejalan dengan temuan Toy et al. [3], bahwa tingkat keberhasilan klasifikasi juga bergantung pada kejelasan polaritas kata dan keunikan konteks. Dalam hal ini, kosakata terkait isu pendidikan seperti “PPDB”, “kuota zonasi”, atau “kebijakan guru” relatif lebih tegas dalam mengindikasikan polaritas tertentu.

Namun demikian, terdapat ruang perbaikan yang cukup signifikan, terutama pada aspek pelabelan dan representasi fitur. Misalnya, dengan menambahkan fitur n-gram atau menggunakan teknik pembobotan yang mempertimbangkan posisi kata dan konteks (*contextual embedding*) seperti *Word2Vec* atau *BERT*, model dapat memperoleh pemahaman yang lebih dalam terhadap struktur kalimat [4]. Penelitian serupa yang menggabungkan pendekatan berbasis frekuensi dan konteks terbukti meningkatkan performa klasifikasi, terutama dalam domain teks sosial yang informal dan padat makna implisit [9].

Terakhir, dalam konteks praktis, temuan ini memberikan kontribusi strategis bagi pemerintah daerah. Analisis berbasis aspek seperti ini memungkinkan identifikasi isu yang paling sensitif di mata publik. Dengan data bahwa aspek ekonomi dan infrastruktur mendominasi perhatian masyarakat, Pemerintah Kota Surabaya dapat memprioritaskan komunikasi kebijakan dan perbaikan pelayanan pada bidang-bidang tersebut secara lebih terfokus. Selain itu, hasil ini juga membuktikan potensi besar penerapan *text mining* dalam perumusan kebijakan berbasis data (*data-driven policy*) di tingkat lokal [5][6].

IV. KESIMPULAN

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kinerja Pemerintah Kota Surabaya melalui pendekatan klasifikasi berbasis aspek menggunakan algoritma *Multinomial Naïve Bayes* dan representasi teks TF-IDF. Data diperoleh dari komentar pengguna media sosial Instagram yang berkaitan dengan empat aspek utama layanan publik, yaitu ekonomi, infrastruktur, kesehatan, dan pendidikan. Hasil analisis menunjukkan bahwa model memberikan performa klasifikasi yang cukup baik, terutama dalam mendeteksi komentar dengan sentimen negatif. Skema pembagian data terbaik diperoleh pada skenario 90:10, dengan akurasi keseluruhan mencapai 88%, dan *f1-score* tertinggi dicapai pada aspek kesehatan (0.90) dan ekonomi (0.93) untuk kelas negatif. Aspek infrastruktur memiliki performa terendah, terutama dalam mengklasifikasikan sentimen positif, yang tercermin dari nilai *recall* sebesar 0.31. Distribusi komentar menunjukkan bahwa topik ekonomi menjadi perhatian utama masyarakat, disusul oleh infrastruktur, kesehatan, dan terakhir pendidikan. Komentar negatif mendominasi hampir semua aspek, mengindikasikan adanya persepsi publik yang kritis terhadap berbagai kebijakan dan layanan pemerintah. Distribusi data yang tidak seimbang antar kelas sentimen dan antar aspek memengaruhi kinerja model dalam mengidentifikasi polaritas

sejara merata. Keberhasilan klasifikasi paling signifikan diperoleh pada aspek-aspek dengan kosakata yang eksplisit dan konsisten. Sementara itu, klasifikasi komentar positif terbukti lebih menantang karena banyaknya ekspresi yang netral atau implisit. Hal ini menunjukkan bahwa meskipun model *Naïve Bayes* dapat menjadi solusi yang efisien untuk analisis opini publik, keberhasilannya sangat bergantung pada kualitas data, fitur representasi, dan strategi pelabelan yang digunakan. Secara keseluruhan, penelitian ini menegaskan bahwa pendekatan *aspect-based sentiment analysis* berbasis algoritma sederhana tetap memiliki potensi kuat dalam membantu pemerintah memahami persepsi publik secara granular. Informasi semacam ini dapat digunakan untuk memetakan isu-isu prioritas yang dirasakan langsung oleh masyarakat, sekaligus sebagai dasar dalam perumusan kebijakan publik yang lebih responsif dan partisipatif.

REFERENSI

- [1] A. N. Rahmanto, “Media Sosial dan Persepsi Publik tentang Good Governance pada Pemerintah Daerah di Solo Raya,” *Jurnal Ilmu Komunikasi*, vol. 20, pp. 88–100, 2022, doi: 10.31315/jik.v20i1.6433.
- [2] A. Jazuli, Widowati, and R. Kusumaningrum, “Aspect-based sentiment analysis on student reviews using the Indo-Bert base model,” *E3S Web of Conferences*, vol. 448, 2023, doi: 10.1051/e3sconf/202344802004.
- [3] I. P. Rahayu, A. Fauzi, and J. Indra, “Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Naive Bayes Dan Support Vector Machine,” *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 2, pp. 296–301, Dec. 2022, doi: 10.30865/json.v4i2.5381.
- [4] W. Zhang, X. Li, Y. Deng, L. Bing, and W. Lam, “A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges,” *ArXiv*, Nov. 2022, doi: <https://doi.org/10.48550/arXiv.2203.01054>.
- [5] Rayuwati, H. Gemasih, and I. Nizar, “IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK MEMPREDIKSI TINGKAT PENYEBARAN COVID,” *JURNAL JURRITEK*, pp. 38–46, Mar. 2022, doi: <https://doi.org/10.55606/jurritek.v1i1.127>.
- [6] T. Rivanie, R. Pebrianto, T. Hidayat, A. Bayhaqy, W. Gata, and H. B. Novitasari, “ANALISIS SENTIMEN TERHADAP KINERJA MENTERI KESEHATAN INDONESIA SELAMA PANDEMI COVID-19,” *Jurnal Informatika*, vol. 21, no. 1, 2021, doi: <https://doi.org/10.30873/ji.v21i1.2864>.
- [7] K. V. S. Toy, Y. A. Sari, and I. Cholissodin, “Analisis Sentimen Twitter menggunakan Metode Naive Bayes dengan Relevance Frequency Feature Selection (Studi Kasus: Opini Masyarakat mengenai Kebijakan New Normal),” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 11, pp. 5068–5074, 2021, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [8] T. S. Utami, “ANALISIS SENTIMEN JUDUL BERITA MENGENAI PEMINDAHAN IBU KOTA NEGARA (IKN) PADA PORTAL BERITA DETIK.COM MENGGUNAKAN SUPPORT VECTOR MACHINE (SVM),” 2023.
- [9] S. Roiqoh, B. Zaman, and Kartono, “Analisis Sentimen Berbasis Aspek Ulasan Aplikasi Mobile JKN dengan Lexicon Based dan Naïve Bayes,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, no. 3, pp. 1582–1592, 2023, doi: 10.30865/mib.v7i3.6194.
- [10] K. A. Siregar, S. Nasution, and P. Nabawy, “Analisis Sentimen Netizen Indonesia Terhadap Kampanye Penggunaan Kecerdasan Buatan Oleh Pemerintah Menggunakan Algoritma Naive Bayes,” *Jurnal Ilmu Komputer dan Informatika*, vol. 1, pp. 112–121, 2025.