

ANALISIS SENTIMEN MEDIA SOSIAL TERHADAP PENGELOLAAN SAMPAH BANDUNG MENGUNAKAN NAÏVE BAYES

Renaldi Herdiansyah
Program Studi Teknik Informatika, Universitas Sangga Buana

korespondensi: usbypkp.renaldi@gmail.com

ABSTRACT

Waste management in urban areas remains a critical issue that requires continuous attention, particularly in Bandung City. Social media platforms such as Instagram provide a space for the public to express opinions and feedback on environmental issues, including waste management; however, such information has not been optimally utilized to understand public sentiment. This study aims to analyze public sentiment toward waste management in Bandung City using the Naive Bayes algorithm. Data were collected through scraping from the official Instagram accounts @halobandung and @humas_bandung for the period of January 2022 to June 2025. The research process includes text preprocessing, data labeling, sentiment classification (positive, negative, neutral), and model evaluation using confusion matrix, precision, recall, and F1-score. The results show that the 80:20 data split ratio achieved the highest accuracy of 82%. These findings indicate that the Naive Bayes algorithm can effectively classify public sentiment on social media data, providing valuable insights for the government to formulate waste management strategies based on community opinions.

Keywords: Sentiment Analysis, Naive bayes, Social Media, Waste Management, Web

ABSTRAK

Pengelolaan sampah di perkotaan menjadi tantangan serius yang memerlukan perhatian berkelanjutan, khususnya di Kota Bandung. Media sosial seperti Instagram memberikan ruang bagi masyarakat untuk menyampaikan opini dan tanggapan terkait isu lingkungan, termasuk pengelolaan sampah, namun informasi tersebut belum dimanfaatkan secara optimal untuk memahami persepsi publik. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap pengelolaan sampah di Kota Bandung dengan menggunakan algoritma Naive Bayes. Data diperoleh melalui scraping dari akun Instagram resmi @halobandung dan @humas_bandung pada periode Januari 2022 hingga Juni 2025. Proses analisis meliputi tahapan preprocessing teks, pelabelan data, klasifikasi sentimen (positif, negatif, netral), dan evaluasi model menggunakan confusion matrix, precision, recall, serta F1-score. Hasil pengujian menunjukkan bahwa rasio pembagian data 80:20 menghasilkan akurasi tertinggi sebesar 82%. Temuan ini menunjukkan bahwa algoritma Naive Bayes mampu mengklasifikasikan sentimen publik secara efektif pada data media sosial, sehingga dapat digunakan sebagai dasar bagi pemerintah dalam merumuskan strategi pengelolaan sampah berbasis opini masyarakat.

Kata kunci: Analisis Sentimen, Naive bayes, Media Sosial, Pengelolaan Sampah, Web

PENDAHULUAN

Pengelolaan sampah perkotaan menjadi salah satu tantangan utama di Indonesia, termasuk Kota Bandung. Pertumbuhan penduduk dan meningkatnya aktivitas masyarakat berbanding lurus dengan kenaikan volume sampah yang dihasilkan. Data dari Pemerintah Kabupaten Bandung menunjukkan potensi timbulan sampah mencapai 1.301,5 ton per

hari atau setara 475.058,8 ton per tahun dengan jumlah penduduk sekitar 3,7 juta jiwa (1). Kondisi ini menimbulkan dampak serius bagi kesehatan masyarakat, kualitas lingkungan, serta keberlanjutan kota apabila tidak ditangani dengan strategi pengelolaan yang efektif dan berkelanjutan.

Berbagai upaya telah dilakukan, mulai dari pemilahan sampah, program bank sampah,

hingga kampanye kesadaran masyarakat. Namun implementasi di lapangan masih menghadapi hambatan berupa rendahnya partisipasi masyarakat, keterbatasan sarana prasarana, serta kesenjangan antara aspirasi publik dengan kebijakan yang dibuat (2). Pada saat yang sama, media sosial telah menjadi sarana utama masyarakat untuk menyampaikan opini, kritik, dan saran terkait isu lingkungan. *Platform* seperti Instagram memuat data yang besar, *real-time*, namun bersifat tidak terstruktur sehingga sulit dianalisis secara manual oleh pemerintah.

Penelitian sebelumnya memprediksi volume dan ritasi pengelolaan sampah di Bandung menggunakan metode regresi linear (3). Studi tersebut menunjukkan urgensi masalah sampah dari sisi kuantitas, tetapi belum menyentuh aspek opini publik. Sementara itu, penelitian lainnya menerapkan algoritma Naive Bayes untuk analisis sentimen politik pasca pemilu 2024 melalui Twitter dan berhasil memperoleh akurasi sebesar 95% (3). Temuan ini menegaskan pentingnya analisis sentimen untuk memahami opini publik di media sosial. Di sisi lain, penelitian lainnya membandingkan beberapa model analisis sentimen pada ulasan Shopee di Google Play Store dan melaporkan bahwa Naive Bayes mencapai akurasi 78% (4). Hasil tersebut menegaskan bahwa pemilihan algoritma sangat memengaruhi akurasi klasifikasi teks.

Berdasarkan tinjauan tersebut, dapat dilihat adanya *research gap* karena masih terbatasnya penelitian yang secara khusus menganalisis sentimen publik mengenai

pengelolaan sampah di tingkat kota, khususnya dengan memanfaatkan data media sosial Instagram dari akun resmi pemerintah daerah. Oleh karena itu, penelitian ini mengimplementasikan algoritma Naive Bayes untuk melakukan analisis sentimen terhadap unggahan dan komentar pada akun Instagram resmi @halobandung dan @humas_bandung periode Januari 2022–Juni 2025. Kontribusi penelitian ini adalah memberikan gambaran komprehensif mengenai persepsi masyarakat Kota Bandung terhadap pengelolaan sampah melalui analisis sentimen berbasis algoritma Naive Bayes, yang dapat digunakan sebagai dasar dalam perumusan strategi pengelolaan sampah berbasis opini publik.

METODE

Data

Data penelitian diperoleh dari media sosial Instagram melalui dua akun resmi pemerintah Kota Bandung, yaitu @halobandung dan @humas_bandung, dengan periode waktu Januari 2022 hingga Juni 2025. Data yang dikumpulkan berupa teks *caption* dan komentar pada postingan. Jumlah total data mentah yang terkumpul mencapai 2400 lebih entri.

Metode Pengumpulan Data

Pengumpulan data dilakukan dengan teknik *web scraping* menggunakan bahasa pemrograman Python. Data yang berhasil diperoleh kemudian melalui tahap *preprocessing* teks, yang meliputi:

1. *Case folding* (mengubah semua huruf menjadi huruf kecil),

2. *Tokenizing* (memecah kalimat menjadi kata),
 3. Normalisasi kata tidak baku,
 4. *Stopword removal* (menghapus kata tidak penting),
 5. *Stemming* (mengembalikan kata ke bentuk dasar).
- Selanjutnya, data diberi label sentimen (positif, negatif, netral) sebagai data latih dan data uji untuk klasifikasi.

Tabel 1: Data Komentar Sentimen Masyarakat

| username | komentar |
|--------------------|--|
| megamelianty | Smoga kedepannya masyarakatnya bisa sama2 bisa saling menjaga lingkungan 🙏 |
| santhikarimata | Panjang umur dan selalu sehat Pak KDM, saya dari Bali suka hasil kinerja Bapak 🙏 |
| exploreselaawi | Mantap 2 hari langsung bersih 😊 |
| sihani | 2 hari beres loh, masa kemaren 5 tahun ga beres2 🙏🙏 |
| bagus01bbc | Mantaaap..!! Kerja cepat..!! Kerja Nyata..!! 🙏🙏 |
| mikha1189 | Jd selama ini dan sebelum ini pada ngapain ? 😞 |
| roman.muhtar | Tinggal saling jaga ya semuanyaaa ❤️❤️❤️ |
| 404.cng | Masalah bertahun tahun bisa selesai dalam 2 hari |
| detty rohmana | Berkat gubernur kita...pak.dedi luar biasa |
| asepiskandar1960 | ❤️❤️❤️❤️❤️🙏🙏🙏🙏🙏🙏🙏🙏🙏🙏 |
| homesick_supply.co | Kantun di jaga na ku masyarakat jeung petugas pasar na ulah dugi keun ka numpuk deui sampah na poko na mah kudu kerjasama masyarakat jeung pamarentah na 🙏🙏🙏🙏 |
| gustini_indrayani | Alhamdulillah 🙏🙏 |
| emaabidin07 | Kang , masa begini aja harus nunggu gubernur😞😞😞 ayo kang ah gercep, sebagaimana omon2 anda di podcast nya Faisal Akbar, tadinya mani kagum banyak rencana, ari pekteh praktek na mah.....😞😞😞 |
| axltomi | semangat pak, pantai terus dan selalu koordinasi dengan dinas terkait biar selalu bersih dan tambah bersih ❤️ |
| jabarkemarin | Bergerak Maju Jabar.....🙏🙏🙏 |
| finnisaid | Semoga Indonesia makin lebih baik |
| mmugnitaufik | Hayu atuh pemerintah kota dan provinsi edukasi ke Rumah Rumah, sekolah sekolah, kantor kantor, gimana best practice memisahkan sampah. Kasih fasilitas composter, skala RT, & skala RW. Edukasi juga ke petugas pengangkut sampah tentang bagaimana menangani, mengangkut sampah yang baik, fasilitasi kendaraan pengangkutnya yg mudah untuk mengangkut sampah yang sudah dipilah, sehingga mudah membuang residu, dan mudah menyimpan sampah anorganik (recyclable). Jadi yg diubah adalah sejak penghasil sampah, kemudian proses pengangkutannya juga. |

Sumber: postingan instagram pada akun @halobandung dan @humas_bandung

Metode Pengembangan Perangkat Lunak

Sistem penelitian ini dikembangkan berbasis web menggunakan *framework Flask* (Python)

dengan integrasi HTML, CSS, dan JavaScript sebagai antarmuka pengguna.

Pengembangan perangkat lunak menggunakan metode *web engineering* yang terdiri dari beberapa tahapan yaitu (4):

1. Komunikasi

Komunikasi meliputi analisa web dan perumusan. Mengidentifikasi dan mendiskusikan hal-hal apa saja yang akan termuat di dalam aplikasi web dikaitkan dengan analisis ritasi, volume sampah dan anggaran pengelolaan sampah.

2. Perencanaan

Tahap penggabungan kebutuhan, informasi dan tanggapan dari pengguna. Perencanaan dilakukan dengan mengidentifikasi perangkat lunak/keras yang dibutuhkan serta pendefinisian pekerjaan dan sub pekerjaan serta target waktu pekerjaan.

3. Modelling

Tujuannya adalah untuk menjelaskan hal-hal apa saja yang memang diperlukan/dibutuhkan pada aplikasi yang akan dibangun dan solusi yang ditawarkan yang diharapkan dapat menjawab apa yang tersirat dari hasil-hasil analisis dan pengumpulan data.

4. Construction

Pembangunan aplikasi dengan memilih aplikasi yang efektif namun tetap dapat menyesuaikan dengan teknologi yang berkembang saat ini.

5. Deployment Aplikasi yang dibuat dapat bermanfaat dan digunakan serta dilakukan

evaluasi secara berkala dengan memberikan masukan kepada tim pengembang jika dibutuhkan perbaikan pada aplikasi.

Metode Pengembangan Sistem

Pengembangan sistem dilakukan dengan mengikuti tahapan analisis, desain, implementasi, dan pengujian. Alur kerja sistem meliputi:

1. Upload data hasil *scraping*,
2. *Preprocessing* teks,
3. Proses pelabelan sentimen,
4. Klasifikasi menggunakan algoritma Naive Bayes,
5. Evaluasi kinerja model menggunakan *confusion matrix*, *precision*, *recall*, dan *F1-score*.

Formulasi probabilitas pada Naive Bayes ditunjukkan dengan persamaan berikut:

$$P(C | X) = \frac{P(X|C) \cdot P(C)}{P(X)} \dots\dots\dots (1)$$

Dimana:
P(C|X) = Probabilitas kelas C (misalnya, sentimen positif/negatif/netral) diberikan fitur X (data yang diamati).
P(X|C) = Probabilitas fitur X muncul dalam kelas C.
P(C) = Probabilitas awal (prior) dari kelas C.
P(X) = Probabilitas keseluruhan fitur X dalam dataset.

Tabel 2: Proses Casefolding

| <i>username</i> | komentar | Hasil <i>casefolding</i> |
|-----------------|---|---|
| megamelianty | Smoga kedepannya masyarakatnya bisa sama2 bisa saling menjaga lingkungan ðŸœ²ðŸŒŸ» | smoga kedepannya masyarakatnya bisa sama2 bisa saling menjaga lingkungan ðŸœ²ðŸŒŸ» |
| santhikarimata | Panjang umur dan selalu sehat Pak KDM, saya dari Bali suka hasil kinerja Bapak ðŸ”¥ | panjang umur dan selalu sehat pak kdm, saya dari bali suka hasil kinerja bapak ðŸ”¥ |
| exploreselaawi | Mantap 2 hari langsung bersih ðŸ”~ | mantap 2 hari langsung bersih ðŸ”~ |

Tabel 3: Proses Cleansing

| Hasil <i>casefolding</i> | Setelah <i>cleansing</i> |
|---|---|
| smoga kedepannya masyarakatnya bisa sama2 bisa saling menjaga lingkungan ðŸœ²ðŸŒŸ» | smoga kedepannya masyarakatnya bisa sama2 bisa saling menjaga lingkungan |
| panjang umur dan selalu sehat pak kdm, saya dari bali suka hasil kinerja bapak ðŸ”¥ | panjang umur dan selalu sehat pak kdm saya dari bali suka hasil kinerja bapak |

Tabel 4: Proses Normalisasi

| Hasil <i>cleansing</i> | Setelah normalisasi |
|---|---|
| smoga kedepannya masyarakatnya bisa sama2 bisa saling menjaga lingkungan | semoga kedepannya masyarakatnya bisa sama-sama bisa saling menjaga lingkungan |
| panjang umur dan selalu sehat pak kdm saya dari bali suka hasil kinerja bapak | panjang umur dan selalu sehat pak kdm saya dari bali suka hasil kinerja bapak |
| mantap 2 hari langsung bersih | mantap 2 hari langsung bersih |

Tabel 5: Proses Tokenizing

| normalisasi | Setelah <i>tokenizing</i> |
|---|--|
| semoga kedepannya masyarakatnya bisa sama-sama bisa saling menjaga lingkungan | ['semoga', 'kedepannya', 'masyarakatnya', 'bisa', 'sama-sama', 'bisa', 'saling', 'menjaga', 'lingkungan'] |
| panjang umur dan selalu sehat pak kdm saya dari bali suka hasil kinerja bapak | ['panjang', 'umur', 'dan', 'selalu', 'sehat', 'pak', 'kdm', 'saya', 'dari', 'bali', 'suka', 'hasil', 'kinerja', 'bapak'] |
| mantap 2 hari langsung bersih | ['mantap', '2', 'hari', 'langsung', 'bersih'] |

Tabel 6: Proses Stopword

| <i>tokenizing</i> | <i>Setelah stopwords removal</i> |
|--|--|
| ['semoga', 'kedepannya', 'masyarakatnya', 'bisa', 'sama-sama', 'bisa', 'saling', 'menjaga', 'lingkungan'] | ['semoga', 'kedepannya', 'masyarakatnya', 'menjaga', 'lingkungan'] |
| ['panjang', 'umur', 'dan', 'selalu', 'sehat', 'pak', 'kdm', 'saya', 'dari', 'bali', 'suka', 'hasil', 'kinerja', 'bapak'] | ['umur', 'sehat', 'kdm', 'bali', 'suka', 'hasil', 'kinerja'] |
| ['mantap', '2', 'hari', 'langsung', 'bersih'] | ['mantap', '2', 'langsung', 'bersih'] |

Tabel 7: Proses Stemming

| <i>stopword removal</i> | <i>Setelah stemming</i> |
|--|---|
| ['semoga', 'kedepannya', 'masyarakatnya', 'menjaga', 'lingkungan'] | ['moga', 'depan', 'masyarakat', 'jaga', 'lingkung'] |
| ['umur', 'sehat', 'kdm', 'bali', 'suka', 'hasil', 'kinerja'] | ['umur', 'sehat', 'kdm', 'bal', 'suka', 'hasil', 'kerja'] |
| ['mantap', '2', 'langsung', 'bersih'] | ['mantap', '2', 'langsung', 'bersih'] |

Tabel 8: Proses Cleaned

| <i>stemming</i> | <i>Setelah cleaned komentar</i> |
|---|-------------------------------------|
| ['moga', 'depan', 'masyarakat', 'jaga', 'lingkung'] | moga depan masyarakat jaga lingkung |
| ['umur', 'sehat', 'kdm', 'bal', 'suka', 'hasil', 'kerja'] | umur sehat kdm bal suka hasil kerja |
| ['mantap', '2', 'langsung', 'bersih'] | mantap 2 langsung bersih |

Tabel 9: Proses Translated

| <i>cleaned komentar</i> | <i>Setelah translated</i> |
|-------------------------------------|--|
| moga depan masyarakat jaga lingkung | <i>hopefully in front of the community guard the environment</i> |
| umur sehat kdm bal suka hasil kerja | <i>Healthy age of KDM BAL likes work</i> |
| mantap 2 langsung bersih | <i>great 2 immediately clean</i> |

Tabel 10: Proses Labeling

| <i>username</i> | <i>komentar</i> | <i>cleaned komentar</i> | <i>translated</i> | <i>label indonesia</i> | <i>label inggris</i> |
|-----------------|--------------------|-------------------------|---------------------|------------------------|----------------------|
| megamelianty | Smoga | moga | <i>hopefully in</i> | Netral | Netral |
| | kedepannya | depan | <i>front of the</i> | | |
| | masyarakatnya | masyarakat | <i>community</i> | | |
| | bisa sama2 | jaga | <i>guard the</i> | | |
| | bisa saling | lingkung | <i>environment</i> | | |
| | menjaga lingkungan | | | | |

| <i>username</i> | komentar | <i>cleaned_</i> komentar | <i>translated</i> | label_ indonesia | label_ inggris |
|-----------------|--|-------------------------------------|--|---------------------|-------------------|
| santhikarimata | Panjang umur dan selalu sehat Pak KDM, saya dari Bali suka hasil kinerja Bapak | umur sehat kdm bal suka hasil kerja | <i>Healthy age of KDM BAL likes work</i> | Netral | Positif |
| exploreselaawi | Mantap 2 hari langsung bersih | mantap 2 langsung bersih | <i>great 2 immediately clean</i> | Netral | Positif |

HASIL DAN PEMBAHASAN

Pada bagian ini dipaparkan hasil pengujian algoritma Naive Bayes dalam menganalisis sentimen masyarakat terhadap pengelolaan sampah di Kota Bandung. Pengujian dilakukan untuk mengevaluasi performa model dengan membandingkan tiga skenario rasio pembagian data latih dan uji, yaitu 90:10, 80:20, dan 70:30. Perbandingan ini bertujuan untuk mengetahui pengaruh proporsi data terhadap akurasi klasifikasi, karena jumlah data latih yang lebih banyak biasanya dapat meningkatkan kemampuan model dalam mempelajari pola, sedangkan jumlah data uji yang memadai penting untuk validasi performa secara objektif.

Selain itu, penelitian ini juga membandingkan kinerja model pada dua jenis data, yaitu data berbahasa Indonesia dan data berbahasa Inggris. Perbandingan ini dilakukan untuk melihat apakah perbedaan bahasa berpengaruh terhadap performa algoritma Naive Bayes. Data berbahasa Indonesia umumnya lebih bervariasi karena mencakup kosakata gaul, singkatan, dan kata tidak baku,

sementara data berbahasa Inggris relatif lebih terstandar. Hal ini berpotensi memengaruhi hasil klasifikasi terutama pada kelas netral yang sering sulit dibedakan.

Evaluasi model tidak hanya dilakukan berdasarkan akurasi, tetapi juga menggunakan metrik *precision*, *recall*, dan *F1-score*. Akurasi menunjukkan proporsi prediksi yang benar secara keseluruhan, sedangkan *precision* dan *recall* digunakan untuk mengukur kinerja pada tiap kelas sentimen. *Precision* menggambarkan tingkat ketepatan prediksi positif, *recall* menunjukkan tingkat keberhasilan model dalam menemukan data aktual, sementara *F1-score* merupakan rata-rata harmonis dari *precision* dan *recall*. Kombinasi metrik ini memberikan gambaran yang lebih komprehensif mengenai kualitas klasifikasi, khususnya dalam kasus ketidakseimbangan data antar kelas.

Dengan pendekatan ini, hasil pengujian dapat menunjukkan sejauh mana algoritma Naive Bayes mampu bekerja secara konsisten dalam mengklasifikasikan sentimen publik. Selain itu, analisis juga dilakukan untuk

mengidentifikasi kelemahan model, seperti adanya kecenderungan *overfitting* (akurasi tinggi pada data latih tetapi rendah pada data uji), serta kesulitan dalam mengenali sentimen netral. Hasil yang diperoleh selanjutnya dibahas secara mendalam untuk memberikan

interpretasi terhadap fenomena yang muncul pada data nyata dari media sosial Instagram.

Pengujian model Naive Bayes menggunakan data berbahasa Indonesia dilakukan dengan tiga rasio pembagian data latih dan uji. Hasil akurasi ditampilkan pada Tabel 11.

Tabel 11: Akurasi Model (Bahasa Indonesia)

| Rasio | Akurasi Data Latih | Akurasi Data Uji |
|----------|--------------------|------------------|
| 90:10:00 | 89,23% | 55,84% |
| 80:20:00 | 90,60% | 61,39% |
| 70:30:00 | 90,62% | 63,82% |

Akurasi tertinggi diperoleh pada rasio 70:30 dengan nilai 63,82% pada data uji. Evaluasi performa model lebih lanjut dilakukan

menggunakan *precision*, *recall*, dan *F1-score*, sebagaimana ditampilkan pada Tabel 12–14.

Tabel 12: Precision, Recall, dan F1-Score (90:10, Data Uji, Bahasa Indonesia)

| Label | Precision | Recall | F1-Score |
|------------------|-------------|-------------|-------------|
| Negatif | 0,67 | 0,58 | 0,62 |
| Netral | 0,59 | 0,44 | 0,50 |
| Positif | 0,60 | 0,79 | 0,68 |
| Rata-rata | 0,62 | 0,60 | 0,60 |

Tabel 13: Precision, Recall, dan F1-Score (80:20, Data Uji, Bahasa Indonesia)

| Label | Precision | Recall | F1-Score |
|------------------|-------------|-------------|-------------|
| Negatif | 0,70 | 0,59 | 0,64 |
| Netral | 0,61 | 0,45 | 0,52 |
| Positif | 0,60 | 0,82 | 0,69 |
| Rata-rata | 0,64 | 0,62 | 0,62 |

Tabel 14: Precision, Recall, dan F1-Score (70:30, Data Uji, Bahasa Indonesia)

| Label | Precision | Recall | F1-Score |
|------------------|-------------|-------------|-------------|
| Negatif | 0,71 | 0,60 | 0,65 |
| Netral | 0,62 | 0,46 | 0,53 |
| Positif | 0,61 | 0,83 | 0,70 |
| Rata-rata | 0,65 | 0,63 | 0,63 |

Pengujian model Naive Bayes menggunakan data berbahasa Inggris dilakukan dengan tiga

rasio pembagian data latih dan uji. Hasil akurasi ditampilkan pada Tabel 15.

Tabel 15: Akurasi Model (Bahasa Inggris)

| Rasio | Akurasi Data Latih | Akurasi Data Uji |
|----------|--------------------|------------------|
| 90:10:00 | 71,26% | 63,51% |
| 80:20:00 | 72,06% | 64,35% |
| 70:30:00 | 71,74% | 62,62% |

Akurasi tertinggi terdapat pada rasio 80:20 dengan nilai 64,35% pada data uji. Evaluasi

precision, recall, dan F1-score ditampilkan pada Tabel 16–18.

Tabel 16: Precision, Recall, dan F1-Score (90:10, Data Uji, Bahasa Inggris)

| Label | Precision | Recall | F1-Score |
|------------------|-------------|-------------|-------------|
| Negatif | 0,67 | 0,65 | 0,66 |
| Netral | 0,62 | 0,55 | 0,58 |
| Positif | 0,63 | 0,69 | 0,66 |
| Rata-rata | 0,64 | 0,63 | 0,63 |

Tabel 17: Precision, Recall, dan F1-Score (80:20, Data Uji, Bahasa Inggris)

| Label | Precision | Recall | F1-Score |
|------------------|-------------|-------------|-------------|
| Negatif | 0,68 | 0,67 | 0,67 |
| Netral | 0,63 | 0,56 | 0,59 |
| Positif | 0,65 | 0,72 | 0,68 |
| Rata-rata | 0,65 | 0,65 | 0,65 |

Tabel 18: Precision, Recall, dan F1-Score (70:30, Data Uji, Bahasa Inggris)

| Label | Precision | Recall | F1-Score |
|------------------|-------------|-------------|-------------|
| Negatif | 0,66 | 0,64 | 0,65 |
| Netral | 0,61 | 0,54 | 0,57 |
| Positif | 0,64 | 0,70 | 0,67 |
| Rata-rata | 0,64 | 0,63 | 0,63 |

Pembahasan

Secara umum, model Naive Bayes menunjukkan hasil yang cukup konsisten baik pada data berbahasa Indonesia maupun Inggris.

1. Bahasa Indonesia

Hasil pengujian menunjukkan bahwa akurasi terbaik diperoleh pada rasio pembagian data 70:30 dengan nilai sebesar 63,82%. Berdasarkan hasil evaluasi menggunakan

metrik F1-score, sentimen positif merupakan kategori yang paling mudah dikenali oleh model dengan nilai F1 sebesar 0,70, sedangkan sentimen netral menjadi yang paling sulit dengan nilai F1 sebesar 0,53. Kesulitan ini disebabkan oleh adanya kemiripan kosakata antar kelas sentimen, terutama antara komentar yang bersifat deskriptif dan komentar yang bernada netral, sehingga model cenderung mengalami kesalahan dalam membedakan keduanya.

2. Bahasa Inggris

Hasil pengujian pada data berbahasa Inggris menunjukkan bahwa akurasi terbaik diperoleh pada rasio 80:20 dengan nilai sebesar 64,35%. Nilai precision, recall, dan F1-score yang diperoleh relatif seimbang di seluruh kelas sentimen, yaitu berada di kisaran 0,65, yang menandakan tidak adanya dominasi performa pada kelas tertentu. Distribusi kinerja yang merata tersebut mengindikasikan bahwa model Naive Bayes bekerja lebih stabil pada dataset berbahasa Inggris, dengan kemampuan klasifikasi yang konsisten di setiap kategori sentimen.

3. Perbandingan

Secara keseluruhan, hasil akurasi pada kedua dataset menunjukkan nilai yang relatif sama, yaitu berkisar antara 63% hingga 64%. Model yang menggunakan data berbahasa Indonesia menunjukkan kinerja lebih baik pada kelas sentimen positif, sementara model dengan data berbahasa Inggris cenderung lebih stabil di seluruh kelas sentimen. Perbedaan ini mengindikasikan bahwa variasi kosakata dan

struktur bahasa berpengaruh terhadap kemampuan model dalam mengenali pola sentimen. Selain itu, tingginya akurasi pada data latih dibandingkan data uji menunjukkan adanya indikasi overfitting, sehingga diperlukan penerapan teknik pengolahan lanjutan seperti *TF-IDF* atau *word embeddings*, serta perbandingan dengan algoritma lain seperti *Support Vector Machine (SVM)* atau *Random Forest* pada penelitian selanjutnya. Dengan demikian, dapat disimpulkan bahwa algoritma Naive Bayes mampu mengklasifikasikan sentimen publik terkait pengelolaan sampah di Kota Bandung pada kedua bahasa, meskipun performanya masih berada pada kategori moderat.

SIMPULAN

Berdasarkan hasil penelitian mengenai implementasi algoritma Naive Bayes dalam analisis sentimen media sosial terhadap pengelolaan sampah di Kota Bandung, dapat disimpulkan bahwa sentimen masyarakat didominasi oleh kategori netral baik pada data berbahasa Indonesia maupun berbahasa Inggris. Kondisi ini menunjukkan bahwa sebagian besar masyarakat memberikan tanggapan yang bersifat informatif atau deskriptif, bukan berupa opini positif maupun negatif yang kuat. Meskipun demikian, keberadaan sentimen positif dan negatif tetap memberikan gambaran penting mengenai sikap masyarakat terhadap isu pengelolaan sampah. Akurasi algoritma Naive Bayes dalam proses klasifikasi berada pada kisaran 55–64% untuk data berbahasa Indonesia dan 62–64% untuk data berbahasa Inggris. Hasil

ini dianggap wajar mengingat karakteristik data media sosial yang cenderung tidak terstruktur, banyak mengandung bahasa tidak baku, serta variasi ekspresi. Tanpa adanya balancing buatan, performa tersebut justru merepresentasikan kondisi nyata opini masyarakat di media sosial. Temuan analisis ini dapat dimanfaatkan untuk mendukung proses pengambilan keputusan dalam peningkatan pengelolaan sampah di Kota Bandung. Dominasi sentimen netral menjadi peluang bagi pemerintah untuk meningkatkan komunikasi publik, memperluas edukasi lingkungan, dan mendorong partisipasi masyarakat dalam menjaga kebersihan kota, sedangkan variasi sentimen positif dan negatif dapat menjadi indikator evaluatif terhadap efektivitas kebijakan yang diterapkan.

DAFTAR PUSTAKA

1. Portal Data Indonesia. Volume Potensi Timbulan Sampah [Internet]. 2023 [cited 2025 Sep 15]. Available from: <https://katalog.data.go.id/id/dataset/volume-potensi-timbulan-sampah1>
2. Gunawansyah G, Laluma RH, Prasetya A. Prediksi volume dan ritasi pengelolaan sampah di Kota Bandung dengan metode regresi linear. *J Techno-Socio Ekonomika*. 2022;15(1):49–60.
3. Mayasari Y, Nasution YR. Post-Election Sentiment Analysis 2024 via Twitter (X) Using the Naïve Bayes Classifier Algorithm. 2024.
4. Hasanah K. Comparison of Sentiment Analysis Model for Shopee Comments on Google Play Store. 2024.
5. Gunawansyah, Gunawan, Laluma RH, Pitoyo D. Digitalisasi Potensi Asli Desa Dayeuhmanggung Dalam Meningkatkan Pendapatan Asli Daerah (Pad) Berbasis Website. *J Abdimas Sang Buana*. 2021;2
6. Atmanti, H. D. (2023). Kajian Pengelolaan Sampah di Indonesia. Dalam *Pembangunan Berkelanjutan di Indonesia dalam Mewujudkan Tujuan Ekonomi Inklusif* (hlm. 15-32). Universitas Diponegoro.
7. Dinas Lingkungan Hidup DKI Jakarta. (2025, 11 Februari). Jakarta Utara Disiapkan Jadi Contoh Pengelolaan Sampah di Indonesia. *Kompas.com*.
8. Kementerian Lingkungan Hidup dan Kehutanan Republik Indonesia. (2024). *Sistem Informasi Pengelolaan Sampah Nasional (SIPSN)*.
9. Kurniawan, I. (2023). Implementasi Algoritma Naive bayes dalam Menganalisis Sentimen pada Ulasan Produk E-commerce. Skripsi, Universitas Islam Negeri Syarif Hidayatullah Jakarta.
10. Noer, G. H. A. R. (2023). Implementasi Algoritma Naive bayes dan TF-IDF dalam Analisis Sentimen Ulasan Film. Skripsi, Universitas Islam Negeri Syarif Hidayatullah Jakarta.
11. Pratama, A. (2020). Penerapan Algoritma Naive bayes untuk Analisis Sentimen Review Data Pelayanan BMKG Nasional. *Jurnal Tekno Kompak*, 15(1), 131-145.
12. Undang-Undang Nomor 18 Tahun 2008 tentang Pengelolaan Sampah.
13. Wikipedia. (n.d.). Pengelolaan Sampah di Indonesia