



Analisis Sentimen Berita Politik Youtube Detik.Com Menggunakan Web Scraping Dan Naive Bayes Classifier

Rofizaidan Maulana¹, Dadang Yusup², Asep Jamaludin³

^{1,2,3}Universitas Singaperbangsa Karawang

Received: 26 Desember 2025

Revised: 31 Desember 2025

Accepted: 02 Januari 2026

Abstract

Dalam era digital, penyampaian berita terkait politik dan konstruksi opini kolektif banyak bergantung pada keberadaan media sosial sebagai saluran utama. Namun, komentar pada berita politik sering mengandung sentimen yang mencerminkan persepsi publik terhadap isu yang berkembang. Kajian ini mengkaji respon afektif pemirsa terhadap pemberitaan politik yang diunggah di situs berbagi video YouTube. dengan menggunakan metode Web Scraping untuk pengumpulan data dan algoritma Naive Bayes Classifier untuk klasifikasi sentimen. Dalam penelitian ini, diterapkan kerangka kerja KDD yang mencakup lima langkah berurutan: memilih data yang relevan, membersihkannya melalui proses pra-pemrosesan, mengubahnya ke dalam format yang sesuai, melakukan eksplorasi data, dan mengevaluasi temuan yang diperoleh. Data diambil dari 1.000 komentar pada lima video berita politik di kanal YouTube Detik.com, yang diklasifikasikan menjadi sentimen positif dan negatif. Sistem penelitian ini menerapkan serangkaian teknik pra-pemrosesan teks yang komprehensif, mulai dari data cleaning, konversi huruf, pembagian token, standarisasi teks, seleksi fitur, hingga stemming, kemudian dilengkapi dengan feature weighting menggunakan pendekatan TF-IDF. Hasil evaluasi menunjukkan bahwa klasifikasi menggunakan Naive Bayes menghasilkan tingkat ketepatan prediksi mencapai 87% pada skema validasi dengan pembagian 90% data latih dan 10% data uji. Namun, klasifikasi sentimen positif memiliki nilai F1- score lebih rendah, menunjukkan kecenderungan model dalam mengenali sentimen negatif. Analisis ini memberikan wawasan bagi pemerintah dan pembuat kebijakan dalam memahami persepsi publik terhadap kebijakan politik serta potensi penyebaran opini negatif di media sosial.

Keywords: Analisis Sentimen, Naive Bayes Classifier, Web Scraping, Berita Politik, YouTube

(*) Corresponding Author:

How to Cite: maulana, rofizaidan, Yusup, D., & Jamaludin, A. (2026). Analisis Sentimen Berita Politik Youtube Detik.Com Menggunakan Web Scraping Dan Naive Bayes Classifier. *Jurnal Ilmiah Wahana Pendidikan*, 12(1.A), 96-107. Retrieved from <https://jurnal.peneliti.net/index.php/JIWP/article/view/12123>.

PENDAHULUAN

Interaksi di media sosial secara substansial memengaruhi konstruksi opini publik dalam konteks isu-isu politik. tetapi juga menjadi sarana penyebaran hoaks yang dapat memicu ketegangan sosial. Contoh nyata adalah kerusuhan pasca-aksi 22 Mei 2019, di mana pemerintah terpaksa membatasi fitur Penggunaan media sosial diarahkan untuk membatasi penyebaran hoaks serta mencegah munculnya konten yang memicu konflik (CNN Indonesia, 2019). Fenomena ini menunjukkan

betapa opini di platform digital dapat berdampak signifikan terhadap stabilitas politik Indonesia.

Menurut Barus (2011:41), politik mencakup segala hal terkait pemerintahan dan negara, termasuk kebijakan publik, dinamika kekuasaan, dan interaksi antara pemimpin dengan masyarakat. Kompleksitas dunia politik menjadikannya subjek yang sering diperdebatkan di ruang publik, termasuk di media sosial. Gambar 1.1 dan 1.2 menunjukkan contoh komentar provokatif dan bermuatan SARA di YouTube Detik.com, yang mempertegas pentingnya analisis lebih mendalam terhadap sentimen publik.

Analisis sentimen (Sentiment Analysis) merupakan Prosedur yang digunakan untuk mengevaluasi isi teks guna menilai kecenderungan sikap, apakah mengarah pada sentimen positif atau negatif. Sistem analisis sentimen modern mampu mencapai akurasi 75–95% tergantung pada jenis data yang digunakan (Nasukawa & Yi, 2003). Dalam konteks Indonesia, pemilihan sumber data yang kredibel sangat krusial. Detik.com, sebagai salah satu platform berita terkemuka, telah terbukti memiliki tingkat akurasi tinggi berdasarkan penelitian Kautsar (2016), yang menganalisis 56 berita dan menemukan konsistensi antara judul, konten, dan sumber.

YouTube, sebagai platform kedua terpopuler di Indonesia (SimilarWeb, 2023), menjadi sumber data potensial untuk memahami opini publik. Penelitian ini mengumpulkan komentar dari channel YouTube Detik.com menggunakan teknik web scraping metode otomatis untuk ekstraksi data dari situs web (Turland, 2010). Pendekatan ini memungkinkan pengumpulan data secara efisien dan akurat.

Algoritma Naïve Bayes dipilih sebagai dasar klasifikasi karena kemampuannya menangani data teks dalam skala besar dengan cepat. Efektivitasnya telah terbukti dalam penelitian sebelumnya, seperti analisis sentimen ulasan film oleh Muhammad Thariq Razaq (2023), di mana kombinasi Naïve Bayes dengan TF-IDF menghasilkan akurasi tinggi. Namun, penerapannya untuk analisis sentimen komentar politik masih perlu dikaji lebih lanjut, mengingat kompleksitas bahasa dan konteks isu politik di Indonesia.

Respons masyarakat terhadap isu di media sosial tidak hanya mencerminkan opini publik, namun juga dapat memberikan tekanan atau pengaruh terhadap proses pengambilan keputusan di tingkat pemerintahan. Sentimen negatif yang masif dapat menjadi indikator ketidakpuasan publik, sementara analisis sentimen berbasis data dapat membantu pemerintah merumuskan strategi komunikasi yang lebih efektif. Dengan demikian, Selain memberikan kontribusi terhadap pengembangan ilmu pengetahuan, temuan dalam penelitian ini juga dapat dimanfaatkan dalam konteks praktis oleh pihak yang berwenang dalam menetapkan kebijakan.

METODE PENELITIAN

Analisis sentimen terhadap tanggapan audiens pada video berita politik di platform YouTube Detik.com dilakukan dengan menerapkan kerangka kerja KDD secara sistematis. Metode ini terdiri dari lima tahapan utama yang akan dijelaskan secara rinci berikut ini.

Tahap pertama penelitian adalah pemilihan data komentar dari channel YouTube Detik.com. Proses akuisisi data dilaksanakan dengan metode scraping web menggunakan library Python khusus untuk ekstraksi konten digital. Komentar yang diambil berasal dari video berita politik yang dipublikasikan sepanjang tahun 2022. Setelah data terkumpul, dilakukan pelabelan manual untuk mengklasifikasikan komentar menjadi dua kategori, yaitu positif dan negatif. Proses pelabelan ini divalidasi oleh ahli bahasa untuk memastikan akurasi.

Data yang telah terkumpul kemudian melalui serangkaian tahap preprocessing untuk mempersiapkan data sebelum analisis. Tahapan ini meliputi:

1. Cleaning

Langkah awal dalam prapemrosesan data melibatkan penghapusan elemen-elemen yang dianggap tidak relevan, seperti karakter khusus, angka, tanda baca, serta emoji, guna memperoleh data teks yang lebih bersih dan terstruktur.

2. Case Folding

Seluruh kata dalam dokumen dikonversi menjadi huruf kecil agar memastikan konsistensi dalam format teks dan menghindari perbedaan makna akibat variasi kapitalisasi.

3. Tokenizing

Kalimat dipisahkan menjadi kata-kata individual menggunakan teknik tokenizing.

4. Filtering

Dilakukan penyaringan kata dengan menghapus stopwords atau unit teks yang keberadaannya tidak memengaruhi makna keseluruhan dalam evaluasi data.

5. Stemming

Kata-kata yang telah difilter kemudian dikembalikan ke bentuk dasarnya menggunakan algoritma stemming.

6. Normalization

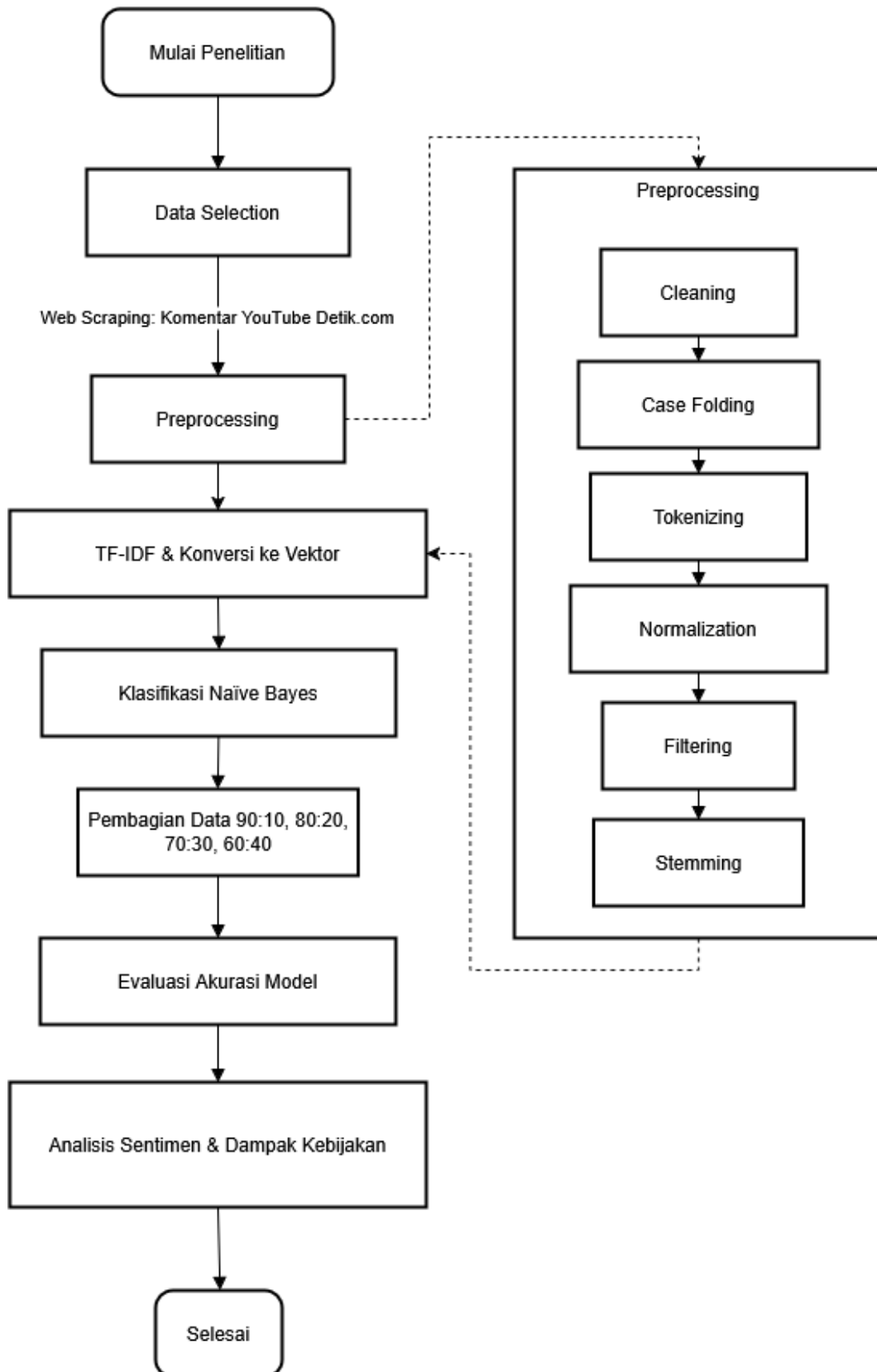
Tahap terakhir adalah normalisasi kata tidak baku menjadi bentuk standar yang sesuai dengan kaidah bahasa Indonesia.

Setelah melalui tahap pra-pemrosesan, dataset mengalami transformasi dengan penerapan teknik TF-IDF yang berfungsi untuk memberikan nilai pembobotan pada setiap term sesuai dengan tingkat relevansi dan distribusinya dalam dokumen. Data kemudian dibagi menjadi beberapa skenario untuk pelatihan dan pengujian model. Ada tiga skenario eksperimen dirancang dengan variasi komposisi dataset:

- Skenario pertama menggunakan 80% data untuk training dan 20% data untuk testing
- Skenario kedua menerapkan proporsi 70% training dan 30% testing
- Skenario ketiga membagi data menjadi 60% untuk training dan 40% testing

Klasifikasi sentimen dilaksanakan dengan memanfaatkan keunggulan algoritma Naïve Bayes dalam menangani karakteristik data teks. Ketiga model kemudian dibandingkan berdasarkan skema pembagian data yang berbeda ini.

Penilaian terhadap model dilakukan melalui analisis Confusion Matrix dan penghitungan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score untuk mengukur tingkat keberhasilan klasifikasi sentimen oleh algoritma Naïve Bayes pada dataset komentar yang berkaitan dengan isu politik.



Gambar 1 Diagram Alur Knowledge Discovery in Databases

HASIL PENELITIAN

Penelitian ini melakukan analisis sentimen terhadap komentar publik di platform YouTube dengan fokus pada konten berita politik dari channel Detik.com. Metodologi yang digunakan mengacu pada Knowledge Discovery in Database (KDD) dengan tahapan yang sistematis mulai dari pengumpulan data hingga interpretasi hasil. Berikut adalah paparan lengkap mengenai proses dan temuan penelitian.

Proses penelitian diawali dengan pengumpulan data komentar dari lima video politik terpilih di channel YouTube Detik.com. Pemilihan video didasarkan pada tingkat interaksi dan relevansi dengan isu politik terkini. Video-video tersebut mencakup berbagai topik seperti dinamika pemilihan presiden 2024, isu koalisi partai politik, hingga analisis mengenai figur-figur politik utama. Total data yang berhasil dikumpulkan mencapai 1000 komentar yang kemudian menjadi basis analisis.

Berikut adalah judul video yang menjadi sumber pengambilan data komentar:

1. #Live! Adu Perspektif: Di Antara Pranowo dan Prabowo (391 komentar)
2. #Mahfud Bicara Islamofobia Hingga Trauma Politik (803 komentar)
3. #Live! Adu Perspektif: Medan Tempur Pasca-putusan MK (776 komentar)
4. #LIVE! Adu Perspektif: Tegang Setelah Gibran Melenggang (681 komentar)
5. #LIVE! Adu Perspektif: Koalisi Besar Vs PDIP, Anies Melenggang? (605 komentar)

Tahap pengumpulan data dilakukan melalui teknik web scraping menggunakan YouTube API v3 dengan bantuan platform Google Colab. Proses ini melibatkan ekstraksi komentar beserta metadata seperti jumlah like. Gambar 4.1 hingga 4.4 dalam penelitian memperlihatkan alur teknis pengambilan data, termasuk implementasi fungsi `video_comment` dengan parameter `video_id` yang memungkinkan ekstraksi data secara spesifik.

Setelah data terkumpul, dilakukan proses seleksi dengan mengambil 200 komentar teratas dari masing-masing video. Pemilihan komentar ini didasarkan pada pertimbangan bahwa komentar dengan like terbanyak cenderung mewakili pandangan yang mendapat resonansi luas di kalangan netizen. Data kemudian disimpan dalam format dataframe menggunakan library `pandas` sebelum diproses lebih lanjut.

Proses pelabelan data dilakukan secara manual dengan bantuan ahli bahasa untuk memastikan validitas klasifikasi sentimen. Dari 1000 komentar yang dianalisis, teridentifikasi 436 komentar bersentimen positif dan 564 komentar bersentimen negatif. Validasi oleh ahli bahasa yang berpengalaman selama lima tahun dalam bidangnya memberikan tingkat kepercayaan yang tinggi terhadap akurasi pelabelan awal ini.

Tabel 1 Contoh Komentar

| No | Komentar | Label | Kode Confusion Matrix |
|----|----------|-------|-----------------------|
| | | | |

| | | | |
|---|---|---------|---|
| 1 | Kita sebagai masyarakat yg baik, supaya saling support tidak usah saling menjatuhkan dan tidak usah saling merendahkan, kita bersaing secara sehat, emang kamu doang yg ingin bisnis kita juga sama2 ingin bisnis | Positif | 1 |
| 2 | Pembawa acara bodoh, tidak bisa mencairkan suasana ,, malah mengadu dasar bab* | Negatif | 0 |
| 3 | Jika ingin negara maju, hambatan harus diminimalisir. Kadang perlu otoriter. Ribut hanya penghambat. | Positif | 1 |
| 4 | Di negeri ini terlalu banyak pejabat yg doyan main drama, sulit dipercaya | Negatif | 0 |
| 5 | Dari awal sdh berbohong, apalagi kl sdh terpilih bisa akan terus membohongi rakyat. Semoga tidak!!! | Negatif | 0 |

Pra-Pemrosesan Data

Tahap pra-pemrosesan data merupakan fase kritis yang menentukan kualitas analisis selanjutnya. Proses ini meliputi enam langkah utama yang dilakukan secara berurutan:

Gambar 2 Contoh komentar

| index | TextKomentar |
|-------|---|
| 0 | bnyk orgn pindah partai itu sdh biasa. PDIP baper... klo yg pindah bkn jokowi ya biasa aja tingkat kepercayaan masy trhdp jokowi besar. Otomatis suara PDIP akn turun. bolh di blng |
| 1 | hasto tololl, presiden nga pernah minta 3 periode, jgn lo framing demi partai lo. Katanya pesantren di Jateng. Klo nga Jokowi dekat dengan Raja MbZ, Dapat apa masyarakat JA |
| 2 | Astungkare PS Saat nya kandangin banten-banteng & PDI-P saat ini maka menyerang gibran dengan isu isu munapik yg. Ingat terpilih tetapi di pilih oleh masyarakat. Jadi bukan dinasti, tidak sedikit anak anak pengu |
| 3 | saya inget dua tokoh PDIP.Pak Trimed,Pak Efendi Simbolon sekarang ngga kelihat,oh iy koruptor yg di tanyakan oleh Pak Mahfud.semoga dgn Pak Mahfud brwdampingan deng |
| 4 | ngakak gw liat Bnag Nusron sama Opung Panda 🤔🤔 Heran yahhh sudah sampai tahap Jabatan yang tinggi , MASIH TIDAK BISA MEMBEDA hanya cara MEMPERMAINKAN LOGIKA dengan LOGIKA . LHA seorang pengusaha ,anggap ecek2 bisa membedakan. LIHAT LAGI DENGAN JELAS , bukan baper yahh MEMPERMAINKAN RASA , KEJUJUR, MORAL DLL. Pemerintah selalu menggo Pemerintah selalu menggaungkan hatihati berita hoax tapi kondisi BERPOLITIK sek PANCASILA tapi kondisi BERPOLITIK sekarang ini???. Nanti kalau saya bilang tent |
| 4 | tindh tapi kondisi BERPOLITIK sekarang ini bagaimana wahhh kamu PRO si ini dan |

Pertama, proses cleaning dilakukan untuk membersihkan data dari unsur-unsur yang tidak relevan seperti emoticon, tanda baca, karakter khusus, dan elemen non-tekstual lainnya. Proses ini menggunakan fungsi khusus yang dirancang untuk menghapus berbagai bentuk noise dalam data. Hasil cleaning menunjukkan peningkatan signifikan dalam keseragaman data yang siap diproses.

Kedua, case folding diterapkan untuk menyelaraskan format teks dengan mengkonversi seluruh huruf kapital menjadi huruf kecil. Penyetaraan ini penting untuk memastikan konsistensi dalam proses analisis selanjutnya. Penggunaan fungsi lower() dari Python terbukti efektif dalam melakukan transformasi ini.

Ketiga, tokenizing memecah komentar menjadi unit-unit kata individual. Proses ini menggunakan word_tokenize dari library NLTK yang mampu mengenali struktur bahasa Indonesia dengan cukup baik. Hasil tokenizing menghasilkan daftar kata yang siap untuk diproses pada tahap berikutnya.

Keempat, normalisasi teks dilakukan untuk menyetarakan variasi penulisan yang tidak baku. Kamus normalisasi yang dikembangkan secara khusus membantu mengubah kata-kata slang, singkatan, atau penulisan tidak standar menjadi bentuk baku yang konsisten. Proses ini secara signifikan mengurangi variasi data yang tidak perlu.

Kelima, filtering menghilangkan stop words atau kata-kata penghubung yang tidak memiliki makna substantif dalam analisis sentimen. Daftar stop words diperluas dengan menambahkan kata-kata spesifik yang sering muncul tetapi tidak relevan dalam konteks analisis.

```

0 [gma, denger, argumen, adi, praysekarang, sad...
1 [penanya, pembawa, acara, berbobot, sperti, ko...
2 [kenceng, jait, baju, tes, urin, bang, ngeri, ...
3 [gass]
4 [bae, laaah]

...
995 [bang, zulfanare, you, okay, generasi, boomers...
996 [budiman, pragmatis, borjuis, wkwkwwk]
997 [hrefhttpsyoutubelurdvrlhttpsyoutubelurdvrla]
998 [pemilihan umum, pilih, partai, bersikap, kons...
999 [judulnya, mestinya, koalisi, vs, koalisi, per...
Name: TextKomentar, Length: 1000, dtype: object
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Unzipping corpora/stopwords.zip.
    
```

Gambar 3 Data Setelah Filtering

Terakhir, stemming menggunakan library Sastrawi berhasil mengurangi kata berimbuhan ke bentuk dasarnya. Proses ini meskipun tidak sempurna, tetapi cukup efektif untuk bahasa Indonesia dan membantu dalam pengelompokan kata yang memiliki akar sama.

Transformasi Data

Pada fase transformasi, data tekstual dikonversi menjadi vektor numerik agar kompatibel dengan pemrosesan algoritma pembelajaran mesin. Studi ini mengimplementasikan skema TF-IDF yang secara efektif mengkuantifikasi signifikansi suatu term relatif terhadap dokumen individual dan seluruh koleksi dokumen.

Empat skema pembagian data yang berbeda (90:10, 80:20, 70:30, dan 60:40) digunakan dalam proses eksperimen untuk menguji ketahanan model

dalam berbagai kondisi. Pembobotan TF-IDF berhasil menangkap karakteristik penting dari data teks dan mengubahnya menjadi vektor fitur yang bermakna.

Tabel 2 Skema Data Pelatihan dan Data Pengujian

| Skema | Data pelatihan | Data pengujian |
|-----------------|----------------------|--------------------|
| Skema 1 (90:10) | 90% (899 data latih) | 10% (100 data uji) |
| Skema 2 (80:20) | 80% (799 data latih) | 20% (200 data uji) |
| Skema 3 (70:30) | 70% (699 data latih) | 30% (300 data uji) |
| Skema 4 (60:40) | 60% (599 data latih) | 40% (400 data uji) |

Data Mining dengan Naïve Bayes

Implementasi algoritma Naïve Bayes Classifier dilakukan menggunakan pendekatan MultinomialNB yang cocok untuk data teks. Keempat skenario pembagian data diuji secara terpisah untuk mengevaluasi performa model dalam berbagai kondisi.

Hasil klasifikasi menunjukkan variasi performa yang menarik. Pada skema 90:10, model mencapai akurasi 87% dengan f1-score 0.93 untuk kelas negatif, tetapi kurang baik dalam mengidentifikasi kelas positif. Pola serupa terlihat pada skema lainnya dengan variasi tingkat akurasi dan presisi yang bergantung pada komposisi data latih dan uji.

Tabel 3 Hasil Klasifikasi

| Skema | Jenis / Label | Jumlah data | <i>Precision</i> | <i>Recall</i> | f-1 score |
|-------|---------------|--------------|------------------|---------------|-----------|
| 90:10 | 0 (negatif) | 87 dari 100 | 0.87 | 1.0 | 0.93 |
| | 1 (Positif) | 13 dari 100 | 0 | 0 | 0 |
| | Macro avg | 100 | 0.435 | 0.5 | 0.465 |
| 80:20 | 0 (negatif) | 159 dari 200 | 0.807 | 1.0 | 0.893 |
| | 1 (Positif) | 41 dari 200 | 1.0 | 0.073 | 0.136 |
| | Macro avg | 200 | 0.903 | 0.536 | 0.514 |
| 70:30 | 0 (negatif) | 236 dari 300 | 0.791 | 1.0 | 0.883 |
| | 1 (Positif) | 64 dari 300 | 0.793 | 0.895 | 0.515 |
| | Macro avg | 300 | 0.895 | 0.515 | 0.472 |
| 60:40 | 0 (negatif) | 312 dari 400 | 0.781 | 1.0 | 0.877 |

| | | | | | |
|--|-------------|-------------|------|-------|-------|
| | 1 (Positif) | 88 dari 400 | 1.0 | 0.011 | 0.022 |
| | Macro avg | 400 | 0.89 | 0.505 | 0.45 |

Analisis mendalam terhadap hasil penelitian mengungkap beberapa temuan penting. Pertama, terdapat ketimpangan signifikan dalam kemampuan model mengklasifikasikan sentimen positif dan negatif. Hal ini kemungkinan

| Positif | | | Negatif | | |
|--------------|-----------|----|--------------|----------|-----|
| Common_words | count | | Common_words | count | |
| 0 | negara | 28 | 0 | orang | 107 |
| 1 | orang | 25 | 1 | politik | 88 |
| 2 | presiden | 25 | 2 | pilih | 79 |
| 3 | pilih | 24 | 3 | partai | 72 |
| 4 | indonesia | 24 | 4 | pimpin | 59 |
| 5 | dalam | 19 | 5 | rakyat | 58 |
| 6 | rakyat | 19 | 6 | presiden | 55 |
| 7 | manusia | 18 | 7 | keluarga | 53 |
| 8 | bangsa | 18 | 8 | negara | 52 |

disebabkan oleh beberapa faktor, termasuk distribusi data yang tidak seimbang dan kompleksitas ekspresi sentimen positif dalam bahasa alami.

Kedua, analisis frekuensi kata mengungkap pola menarik dalam ekspresi sentimen masyarakat. Kata "negara" mendominasi komentar positif, seringkali dalam konteks harapan dan aspirasi. Sebaliknya, kata "orang" banyak muncul dalam komentar negatif, menunjukkan kecenderungan masyarakat untuk mengkritik aktor atau oknum tertentu daripada sistem secara keseluruhan.

Temuan ini memiliki implikasi penting bagi pemahaman tentang dinamika opini publik di ruang digital. Ekspresi sentimen positif cenderung lebih abstrak dan berorientasi pada konsep-konsep besar seperti negara, sementara sentimen negatif lebih sering terfokus pada individu atau kelompok tertentu.

Performa Model

Evaluasi menyeluruh terhadap performa model menunjukkan bahwa skema 90:10 memberikan hasil terbaik dalam hal akurasi keseluruhan (87%), meskipun dengan keterbatasan dalam mengidentifikasi sentimen positif. Skenario 70:30 menawarkan keseimbangan yang lebih baik antara pengenalan sentimen positif dan negatif, dengan f1-score 0.515 untuk kelas positif.

Tabel 4 Skema dan Akurasi

| Skema | Akurasi |
|-------|-------------|
| 90:10 | 0.87 |

| | |
|-------|-------|
| 80:20 | 0.81 |
| 70:30 | 0.793 |
| 60:40 | 0.782 |

Hasil ini mengindikasikan bahwa analisis sentimen untuk konten politik di Indonesia, komposisi data latih yang lebih besar tidak selalu menghasilkan performa yang lebih baik. Faktor kualitas data dan representasi yang seimbang antar kelas mungkin lebih penting daripada sekadar jumlah data.

Implikasi Praktis

Temuan penelitian ini memberikan wawasan berharga bagi berbagai pemangku kepentingan. Bagi media, memahami pola sentimen penonton dapat membantu dalam menyusun konten yang lebih responsif. Bagi aktor politik, analisis ini mengungkap bagaimana publik merespons berbagai isu dan narasi politik.

Keterbatasan penelitian terutama terletak pada cakupan data yang hanya berasal dari satu platform media. Penelitian lanjutan dapat memperluas cakupan sumber data dan mempertimbangkan faktor-faktor tambahan seperti intensitas emosi dalam ekspresi sentimen.

Dari segi metodologi, penelitian ini menunjukkan bahwa pendekatan Naïve Bayes masih relevan untuk analisis sentimen bahasa Indonesia, meskipun dengan beberapa keterbatasan. Pengembangan kamus istilah politik yang lebih komprehensif dan penyempurnaan teknik pra-pemrosesan dapat meningkatkan performa model di masa depan.

KESIMPULAN

Berdasarkan hasil penelitian tentang analisis sentimen berita politik di YouTube *Detik.com* menggunakan *web scraping* dan algoritma *Naive Bayes Classifier*, dapat disimpulkan hal-hal berikut:

1. Akurasi Model Algoritma Naïve Bayes mencapai akurasi tertinggi 87% pada skema pembagian data 90:10 (90% data latih, 10% data uji). Namun, model cenderung lebih efektif mengidentifikasi sentimen negatif (F1-score 0,93) dibandingkan sentimen positif (F1-score 0). Hal ini menunjukkan bahwa ekspresi sentimen negatif dalam komentar politik memiliki pola yang lebih konsisten, sementara sentimen positif lebih variatif dan sulit diklasifikasikan.
2. Pola Sentimen Publik Analisis frekuensi kata mengungkap bahwa: Komentar positif didominasi kata "negara", yang sering muncul dalam konteks harapan dan aspirasi terhadap kebijakan politik. Komentar negatif banyak menggunakan kata "orang", mengindikasikan kecenderungan masyarakat untuk mengkritik figur/oknum tertentu daripada sistem secara keseluruhan.
3. Implikasi Metodologis Ketimpangan data (564 komentar negatif vs. 436 positif) memengaruhi performa model. Penelitian selanjutnya perlu menyeimbangkan dataset atau menggunakan teknik oversampling/undersampling. Preprocessing (termasuk normalisasi dan stemming) terbukti krusial dalam meningkatkan kualitas analisis untuk bahasa Indonesia yang bersifat kontekstual. TF-IDF efektif sebagai

pembobotan, tetapi perlu dikombinasikan dengan metode lain (seperti word embedding) untuk menangkap nuansa bahasa politik yang kompleks.

4. Implikasi Praktis Bagi pemerintah dan politisi, temuan ini dapat menjadi acuan dalam merespons opini publik, terutama dalam mengidentifikasi isu-isu yang rentan memicu sentimen negatif. Bagi media, analisis sentimen dapat digunakan untuk memantau reaksi audiens dan menyusun konten yang lebih responsif. Bagi peneliti, pendekatan ini dapat dikembangkan dengan menambahkan analisis emosi (beyond positif/negatif) atau menggabungkan multi-platform (Twitter, Facebook) untuk cakupan lebih luas.

SARAN

Untuk pengembangan penelitian selanjutnya, berikut beberapa saran yang dapat dipertimbangkan:

1. Mencoba untuk mengimplementasi algoritma selain naïve bayes untuk mengetahui perbandingan hasil dari masing-masing algoritma dan untuk mengetahui algoritma mana yang paling optimal untuk digunakan dalam proses *Sentiment Analysis* khususnya terkait topik berita politik. Adapun algoritma yang dapat digunakan dan dibandingkan diantaranya adalah *KNN* dan *support vector machine*.
2. Menambahkan lebih banyak data dan variasi sumber (misalnya, dari platform media sosial lain seperti Twitter atau Instagram) untuk memperluas cakupan analisis. Mempertimbangkan klasifikasi sentimen menjadi tiga kategori (positif, negatif, netral) untuk mendapatkan pemahaman yang lebih mendalam.
3. Menggunakan teknik *cross-validation* untuk memastikan keandalan model. Melibatkan lebih banyak ahli bahasa atau menggunakan leksikon sentimen yang lebih komprehensif untuk pelabelan data.

REFERENSI

- Alexander, A., & Firza, M. H. H. (2023). Analisis Kesalahan Ejaan dan Tanda Baca Pada Salah Satu Surat Kabar. *Jurnal Ilmiah Dan Karya Mahasiswa*, 1(1), 24–28.
- Alfianistiawati, R. (2021). Konstruksi media massa dalam pembentukan stigma masyarakat mengenai covid-19. *Jurnal Ilmu Komunikasi Acta Diurna*, 17(2).
- Alwasi'a, A. (2020). *Analisis Sentimen pada review Aplikasi Berita Online Menggunakan Metode Maximum Entropy (Studi Kasus: Review Detikcom pada Google Play 2019)*.
- Arhami, M., Kom, M., & Muhammad Nasir, S. T. (2020). *Data Mining-Algoritma dan Implementasi*. Penerbit Andi.
- Darmawan, T. D. (2022). Analisis Sentimen Review Pelanggan E-Commerce di Indonesia Menggunakan Algoritma *Naive Bayes Classifier*. *Universitas Dinamika*.
- Hardjo, S. (2019). Hubungan Antara Persepsi Terhadap Berita Politik di Televisi dengan Kecemasan Ibu Rumah Tangga Akan Tindak kejahatan. *Universitas Medan Area*.

- Herdhianto, A. (2020). *Sentiment Analysis* menggunakan Naïve Bayes Classifier (NBC) PADA tweet tentang zakat. *Fakultas Sains Dan Teknologi Universitas Islam Negeri Syarif*.
- Hikmawan, S., Pardamean, A., & Khasanah, S. N. (2020). Sentimen Analisis Publik Terhadap Joko Widodo terhadap wabah Covid-19 menggunakan Metode Machine Learning. *Jurnal Kajian Ilmiah*, 20(2).
- Mahariani, A., & Nurmalasari, D. (2022). Distribution and Classification of Community Service Topics Based on the Results of Extracting Information from Proposal Documents Using *Text Mining* With Naive Bayes Algorithm (Case Study: Polytechnic Caltex Riau). *Jurnal Aksara Komputer Terapan*, 11.
- Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan *Confusion Matrix* Pada Analisis Sentimen Berbasis Teks Pada Twitter. *J-SAKTI (Jurnal Sains Komputer Dan Informatika)*, 5(2), 697–711.
- Ruslim, K. I., Adikara, P. P., & Indriati, I. (2019). Analisis Sentimen Pada Ulasan Aplikasi Mobile Banking Menggunakan Metode Support Vector Machine dan Lexicon Based Features. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 3(7), 6694–6702.
- Singgalen, Y. A. (2023). Analisis Sentimen Wisatawan terhadap Kualitas Layanan Hotel dan Resort di Lombok Menggunakan SERVQUAL dan CRISP-DM. *Building of Informatics, Technology and Science (BITS)*, 4(4), 1870–1882.