ANALISIS SENTIMEN *TWEET* MASYARAKAT PADA ISU INDONESIA GELAP DI MEDIA SOSIAL X MENGGUNAKAN ALGORITMA *NAÏVE BAYES* DAN METODE *COUNTVECTORIZER*

Ramadhani, Taufik Hidayat, Sukisno
Teknik Informatika, Universitas Islam Syekh-Yusuf
Jl. Maulana Yusuf No.10 Babakan, Kec. Tangerang, Kota Tangerang, Banten 15118
Ramadhanigb19@gmail.com

ABSTRAK

Platform media sosial X telah menjadi platform utama bagi masyarakat untuk menyuarakan pendapat mereka, salah satunya adalah isu "Indonesia Gelap" (Indonesia Gelap), yang mencerminkan ketidakpuasan publik terhadap keputusan pemerintah salah satunya yaitu pemangkasan anggaran pendidikan dan revisi uu minerba. Studi ini bertujuan untuk menganalisis sentimen publik terhadap isu tersebut dengan mengklasifikasikan tweet ke dalam kategori positif, negatif, dan netral. Data dikumpulkan dari X selama periode 1–28 Februari 2025, menghasilkan 2.534 tweet bersih setelah prapemrosesan. Studi ini menerapkan algoritma Naïve Bayes untuk klasifikasi dengan ekstraksi fitur menggunakan CountVectorizer. Hasil menunjukkan bahwa sentimen publik didominasi oleh opini negatif (87,7%), diikuti oleh sentimen positif (6,8%) dan netral (5,5%). Evaluasi model menggunakan matriks kebingungan menunjukkan tingkat akurasi keseluruhan sebesar 86%. Model ini berkinerja sangat baik dalam mengklasifikasikan sentimen negatif dengan skor F1 sebesar 93%, namun kinerjanya lebih rendah untuk kelas positif dan netral akibat ketidakseimbangan data yang signifikan. Visualisasi data melalui WordCloud memperkuat temuan ini, menunjukkan dominasi kata-kata dengan nada ketidakpuasan seperti 'orang', 'negara', dan 'darurat' dalam sentimen negatif. Studi ini mengonfirmasi keefektifan Naïve Bayes dalam menganalisis tren sentimen pada data media sosial.

Kata kunci : *Analisis Sentimen, X, Naïve Bayes, Countvectorizer, Indonesia Gelap*

1. PENDAHULUAN

Pesatnya perkembangan teknologi komunikasi telah melahirkan berbagai platform digital yang menjadi sarana masyarakat untuk berinteraksi dan berbagi informasi [1].

Media sosial, khususnya platform X (sebelumnya Twitter), telah bertransformasi menjadi ruang publik virtual utama di Indonesia, dengan total 24 juta pengguna pada tahun 2023 [2].

Platform ini memungkinkan penyebaran opini dan informasi secara cepat melalui unggahan singkat yang disebut *tweet* [3].

Salah satu isu yang menjadi sorotan dan sempat menjadi *trending topic* adalah tagar "Indonesia Gelap", yang digunakan secara luas oleh masyarakat untuk menyuarakan ketidakpuasan dan keresahan terhadap berbagai kebijakan pemerintah serta kondisi sosial-politik di tanah air [4].

Fenomena ini menunjukkan betapa pentingnya media sosial sebagai cerminan opini publik.

Meskipun tagar "Indonesia Gelap" masif digunakan, pemahaman mengenai persepsi publik yang terkandung di dalamnya apakah cenderung positif, negatif, atau netral belum terpetakan secara kuantitatif. Kurangnya analisis mendalam terhadap sentimen di balik isu ini menjadi sebuah celah, padahal pemahaman tersebut sangat krusial bagi pemerintah untuk mengevaluasi kebijakan, serta bagi akademisi dan media untuk memahami dinamika sosial yang sedang terjadi [5].

Oleh karena itu, muncul kebutuhan untuk melakukan analisis sentimen guna mengklasifikasikan dan mengukur pandangan masyarakat terhadap isu "Indonesia Gelap" secara sistematis. Penelitian ini memiliki dua tujuan utama. Pertama, mengklasifikasikan sentimen masyarakat terhadap isu "Indonesia Gelap" di media sosial X ke dalam kategori positif, negatif, dan netral untuk mengetahui distribusi opini publik. Kedua, mengukur performa dan efektivitas algoritma *Naïve Bayes Classifier* yang dikombinasikan dengan metode ekstraksi fitur *CountVectorizer* dalam menangani analisis sentimen pada data teks berbahasa Indonesia yang bersifat dinamis dan kontekstual.

Untuk mencapai tujuan tersebut, penelitian ini akan dilaksanakan melalui beberapa tahapan sistematis. Tahapan dimulai dengan pengumpulan data (*crawling*) *tweet* menggunakan kata kunci "Indonesia Gelap". Selanjutnya, data mentah akan melalui serangkaian proses pra-pemrosesan teks untuk membersihkan *noise*. Fitur dari teks bersih akan diekstraksi menggunakan *CountVectorizer* sebelum akhirnya diklasifikasikan dengan algoritma *Naïve Bayes*. Kinerja model klasifikasi akan dievaluasi menggunakan metrik *confusion matrix* untuk mengukur akurasi, presisi, dan *recall*. Hasil analisis akan divisualisasikan untuk memberikan gambaran yang komprehensif mengenai sentimen publik. Pendekatan ini dipilih karena *Naïve Bayes* dikenal efisien dan memiliki akurasi yang solid untuk klasifikasi teks [6].

Sementara *CountVectorizer* efektif dalam mengubah data teks menjadi format numerik yang dapat diolah model [7].

Beberapa penelitian sebelumnya telah menerapkan analisis sentimen pada berbagai isu. Penelitian oleh [8] menggunakan Naïve Bayes untuk menganalisis sentimen penutupan TikTok Shop dan memperoleh akurasi 89.23%.

Penelitian serupa dilakukan oleh [9] juga menerapkan Naïve Bayes untuk menganalisis ulasan aplikasi TikTok Shop dan memperoleh akurasi 87,6%, menunjukkan konsistensi metode ini pada data ulasan produk. Di sisi lain, penelitian oleh [10] menggunakan algoritma Support Vector Machine (SVM) untuk menganalisis sentimen terkait konflik Palestina-Israel menunjukkan sentimen positif dominan (81 post). Akun centang biru lebih aktif (78 akun) dibandingkan non-centang biru (72 akun), sementara itu penelitian [11] menggunakan K-Nearest Neighbors (KNN) pada opini tentang ChatGPT menggunakan dataset 1000 opini pengguna media sosial X yang dikumpulkan, bahwa 853 opini bersifat positif dan 147 opini bersifat negatif terhadap ChatGPT dengan akurasi 84.80%.

Studi oleh [12] kembali menggunakan Naïve Bayes untuk sentimen *cryptocurrency* dan mendapatkan akurasi 71.98%. Berbeda dari penelitian-penelitian tersebut yang berfokus pada produk komersial atau isu internasional, penelitian ini berfokus pada isu politik domestik "Indonesia Gelap" dan secara spesifik menerapkan kombinasi algoritma *Naïve Bayes* dengan metode ekstraksi fitur *CountVectorizer*. *Naïve Bayes* dipilih karena efisiensi komputasinya dan kemampuannya dalam klasifikasi teks berbasis probabilitas [6], sementara *CountVectorizer* digunakan untuk mengonversi data teks menjadi bentuk numerik yang dapat diproses oleh model *machine learning* [7].

Tujuan penelitian ini adalah untuk mengetahui kinerja algoritma *Naïve Bayes* dalam mengklasifikasikan sentimen pada isu "Indonesia Gelap" dan menganalisis bagaimana distribusi sentimen publik terhadap isu tersebut.

Penelitian ini dapat memberikan manfaat bagi berbagai pihak. Pemerintah dapat menggunakan hasil analisis ini untuk mengidentifikasi keresahan masyarakat dan merancang kebijakan yang lebih responsif terhadap aspirasi publik. Media dan akademisi juga dapat menjadikannya sebagai bahan kajian dalam memahami dinamika opini publik di media sosial. Berdasarkan latar belakang yang telah diuraikan, maka penelitian ini mengambil judul "Analisis Sentimen Tweet Masyarakat Pada Isu Indonesia Gelap Menggunakan *Naïve Bayes* dan *Countvectorizer*".

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen merupakan cabang ilmu data dan pengolahan bahasa alami yang bertujuan untuk mengenali dan mengekstrak data subjektif, seperti pendapat dan emosi yang ada dalam teks [13].

Pemanfaatan analisis sentimen dapat mendukung pemahaman mengenai kecenderungan komentar yang diberikan oleh pengguna *platform* media sosial, dengan mengklasifikasikan apakah komentar tersebut berorientasi pada sentimen positif atau negatif [14].

2.2. Naïve Bayes

Algoritma Naïve Bayes adalah metode klasifikasi probabilistik berdasarkan Teorema Bayes dengan asumsi independensi (naif) antar fitur. Meskipun asumsinya sederhana, algoritma ini sangat efektif dan efisien untuk klasifikasi teks [15].

Probabilitas posterior dihitung menggunakan rumus berikut:

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (1)$$

Di mana:

- P(c|x) adalah probabilitas kelas *c* diberikan data *x*.
- P(x|c) adalah probabilitas data *x* diberikan kelas *c*.
- P(c) adalah probabilitas prior dari kelas *c*.
- P(x) adalah probabilitas prior dari data *x*.

2.3. Conutvectorizer

Metode *CountVectorizer* dalam penelitian ini berfungsi mengubah korpus teks pada *dataset* menjadi matriks berdasarkan jumlah token kata yang ada. Setiap kata unik dalam teks direpresentasikan sebagai fitur, dan nilai dari setiap fitur adalah jumlah kemunculannya dalam dokumen [16].

Matrik yang dihasilkan mencerminkan frekuensi kata dalam setiap dokumen, di mana baris matriks mewakili dokumen dan kolom mewakili kata. Setiap sel matriks berisi jumlah kemunculan kata dalam dokumen terkait [17].

2.4. Evalusi Model

Kinerja model klasifikasi dievaluasi menggunakan *confusion matrix*, yang membandingkan nilai prediksi dengan nilai aktual. Dari matriks ini, metrik seperti *Accuracy*, *Precision*, *Recall*, dan *F1-Score* dihitung untuk mengukur performa model [18].

Tabel 1. Confusion Matrik

	TRUE	**FALSE**
TRUE	TP	FP
FALSE	FN	TN

Pada *Counfusion matriks* mempunyai empat hasil yang dapat diperoleh dari data pada tabel diatas yaitu: akurasi, presisi, *recall*, *F1-Score*. Penjelasan lebih lanjut dibawah ini.

a. *Accuracy*
Jumlah prediksi yang benar terhadap total prediksi yang dibuat model menggunakan rumus:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} X\ 100\% \quad (2)$$

b. *Precision*
Akurasi dari prediksi positif suatu model, menggunakan rumus:

$$Precision = \frac{TP}{TP + FP} \; X \; 100\% \quad (3)$$

c. *Recall*
Ketika model dapat menangkap semua sampel positif yang benar.

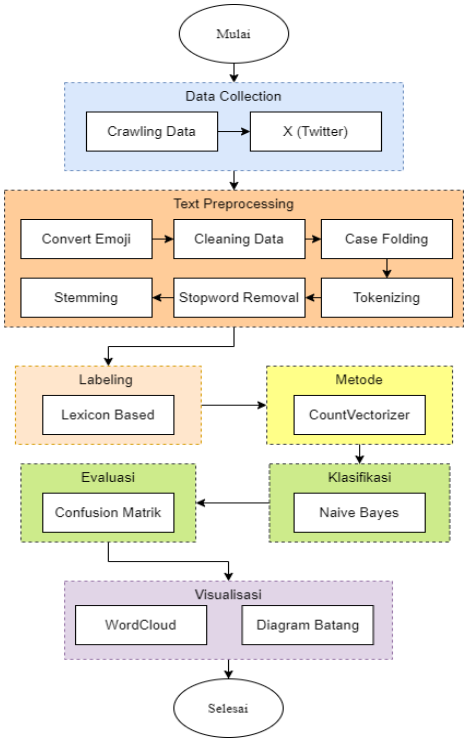
$$Recall = \frac{TP}{TP + FN} \; X \; 100\% \quad (4)$$

d. *F1-Score*
Rata-rata dari *precision* dan *recall.*

$$F1 - Score = \frac{(\,2 \; x \; Recall \; x \; Precision\,)}{Recall + Precision} \; X \; 100\% \quad (5)$$

3. METODE PENELITIAN

Alur penelitian yang disusun oleh peneliti bertujuan untuk menyederhanakan pelaksanaan riset ini. Kerangka kerja penelitian disajikan secara visual pada Gambar 1.



Gambar 1. Alur Penelitian

Penelitian ini menggunakan alur kerja yang sistematis, dimulai dari pengumpulan data hingga visualisasi hasil.

a. Crawling Dataset
Data yang digunakan dalam penelitian ini berasal dari *tweet* di platform media sosial X yang membahas isu “Indonesia Gelap”, dikumpulkan pada periode 1 Februari 2025 hingga 28 Februari 2025, sejalan dengan dinamika sosial-politik yang terjadi pada awal tahun 2025. Pengambilan data dilakukan menggunakan *Tweet Harvest*, sebuah alat crawling yang memungkinkan ekstraksi tweet berdasarkan kata kunci spesifik, yaitu “Indonesia Gelap” serta frasa lain yang relevan dengan isu tersebut. Data tweet yang terkumpul disimpan dalam format .csv untuk analisis lebih lanjut.

b. Teks Preprocessing
Data mentah yang diperoleh masih mengandung banyak *noise*. Oleh karena itu, dilakukan serangkaian tahap *preprocessing* untuk membersihkan data teks. Tahapan tersebut meliputi:
- Convert Emoji: Mengubah emoji menjadi representasi teksnya.
- Cleaning Data: Menghapus URL, *mention*, tagar, tanda baca, dan karakter yang tidak relevan.
- Case Folding: Mengubah semua teks menjadi huruf kecil.
- Tokenizing: Memecah kalimat menjadi token kata individual.
- Stopword Removal: Menghapus kata-kata umum yang tidak memiliki makna signifikan (misalnya, "di", "dan", "yang").
- Stemming: Mengubah kata ke bentuk dasarnya (misalnya, "memakan" menjadi "makan"). Setelah proses ini, diperoleh 2.534 *tweet* bersih.

c. Pelabelan Dataset
Pelabelan sentimen dilakukan secara otomatis menggunakan pendekatan berbasis leksikon, yaitu kamus *InSet Lexicon* untuk Bahasa Indonesia. Setiap *tweet* diberi skor polaritas, kemudian diklasifikasikan sebagai positif (skor > 0), negatif (skor < 0), atau netral (skor = 0).

d. Eksktrasi Fitur dan Klasifikasi
Metode *CountVectorizer* digunakan untuk mengubah data teks yang telah bersih menjadi matriks fitur numerik. Dataset kemudian dibagi menjadi 80% data latih dan 20% data uji. Model klasifikasi dibangun menggunakan algoritma Naïve Bayes.

e. Evaluasi Dan Visualisasi
Kinerja model dievaluasi menggunakan *confusion matrix* dan metrik turunannya. Hasil analisis kemudian divisualisasikan menggunakan diagram batang untuk distribusi sentimen dan *WordCloud* untuk menampilkan kata-kata yang paling sering muncul di setiap kategori sentimen.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Dataset

Proses pengumpulan data dari media sosial X menggunakan kata kunci "Indonesia Gelap" selama periode 1 hingga 28 Februari 2025. Data yang akan diambil adalah *date*, *username* dan *tweet* berhasil mengumpulkan sebanyak 3.392 tweet mentah, berikut adalah Contoh lima baris dari dataset.

Tabel 2. Hasil Crawling

Date	User name	Tweet
Fri Feb 21 10:29:19 +0000 2025	Saham_ fess	Beneran indonesia gelap sih ini wkwkwkwk shm! https://t.co/AGOjlgAqAA
Sun Feb 09 02:42:23 +0000 2025	Tifflylu vh	Indonesia makin hari makin gelap

4.2. Hasil Teks Preprocessing

Data mentah yang telah dikumpulkan kemudian melalui serangkaian tahap *preprocessing* untuk membersihkan data dari *noise* yang dapat mengganggu proses analisis. Tahapan ini meliputi konversi emoji, *cleaning* (menghapus URL, mention, tagar, dan tanda baca), *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Dari 3.392 tweet mentah, setelah proses pembersihan, diperoleh 2.534 tweet yang bersih dan siap untuk dianalisis lebih lanjut. Contoh perbandingan data sebelum dan sesudah *preprocessing* disajikan pada Tabel 3.

Tabel 3. Tweet sesudah dan sebelum

Tweet asli	Hasil Preprocessing
Beneran indonesia gelap sih ini wkwkwkwk shm! https://t.co/AGOjlgAqAA	beneran indonesia gelap sih wkwkwkwk shm
Indonesia makin hari makin gelap	indonesia makin hari makin gelap

4.3. Labeling Sentimen

Setelah data bersih, Pelabelan sentimen dilakukan dengan menggunakan teknik yang didasarkan pada kamus *lexikon*. Setiap tweet dikelompokkan ke dalam tiga jenis positif, negatif, atau netral. Hasil dari proses pelabelan mengindikasikan bahwa persepsi masyarakat mengenai isu "Indonesia Gelap" lebih banyak didominasi oleh sentimen negatif. Distribusi hasil pelabelan dapat dilihat pada Tabel 4. Pada Tabel 5 bisa dilihat hasil dari labeling dataset.

Tabel 4. Distribusi Hasil Pelabelan

Kategori	Jumlah Tweet	Persentase
Negatif	2.223	87.7%
Positif	173	6.8%
Netral	138	5.5%

Tabel 5. Hasil Labelling dataet

Tweet	Skor	Kategori
bendera indonesia indonesian army kopassus using dark camo	-3	Negatif
indonesia gebrak jungwoo gondrong anjayyyyy	0	Netral
indonesia gelap presiden tukang curhat omon omon omong kasar	-13	Negatif

4.4. Countvectorizer

Setelah data teks dibersihkan dan dilabeli, langkah selanjutnya adalah mengonversi teks menjadi bentuk angka yang dapat diolah oleh algoritma pembelajaran mesin. Proses ini disebut ekstraksi fitur, dan pada penelitian ini digunakan metode *CountVectorizer*. Hasil dari proses ini adalah sebuah matriks dokumen-kata (*document-term matrix*), yang kemudian digunakan sebagai input untuk melatih dan menguji model klasifikasi *Naïve Bayes.* Tabel 6 merupakan contoh hasil yang didapatkan setelah dataset melalui proses *countvectorizer* sebagai berikut:

Tabel 6. Hasil Countvectorizer

No	Abangku	…..	Zolim	Zon	zuhud
1	0	…..	0	0	0
2	0	…..	0	0	0
3	0	…..	0	0	0
4	0	…..	0	0	0
….	….	…..	….	….	….
1332	0	…..	0	0	0

4.5. Evaluasi Model

Setelah data tweet berhasil diubah menjadi representasi numerik menggunakan metode CountVectorizer, tahap selanjutnya dalam penelitian ini adalah melakukan proses klasifikasi. Klasifikasi bertujuan untuk mengklasifikasikan data tweet ke dalam tipe sentimen yang spesifik, yaitu sentimen yang bersifat positif, netral, atau negatif.

Algoritma yang digunakan untuk melakukan klasifikasi adalah metode *naïve bayes*. Sebelum melakukan klasifikasi, dataset akan dibagi menjadi dua bagian, yaitu data pelatihan dan data pengujian, dengan proporsi 80% untuk data pelatihan dan 20% untuk data pengujian. Proses ini bertujuan untuk mengukur efektivitas model terhadap data yang belum pernah dilihat sebelumnya.

Tabel 7. Hasil Evaluasi Model

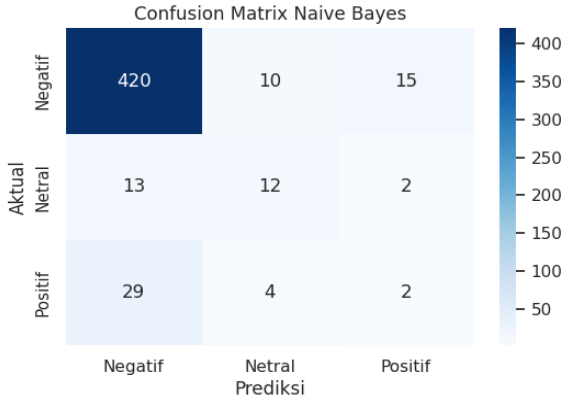
Kelas	Hasil Pengujian		
	Presisi	Recall	F1-Score
Negatif	91%	94%	93%
Netral	46%	44%	45%
Positif	11%	6%	7%
Akurasi			86%

Model klasifikasi sentimen menggunakan algoritma Naïve Bayes menunjukkan performa yang bervariasi antar kelas. Untuk kelas negatif, model mencapai presisi 91%, recall 94%, dan F1-score 93%, mengindikasikan kemampuan deteksi komentar negatif yang sangat baik. Pada kelas netral, performa lebih rendah dengan presisi 46%, recall 44%, dan F1-score 45%, kemungkinan akibat keterbatasan data atau ketidakseimbangan kelas. Untuk kelas positif, model menunjukkan performa terlemah dengan presisi 11%, recall 6%, dan F1-score 7%, dipengaruhi oleh ketidakseimbangan data yang signifikan.

Secara keseluruhan, model mencapai akurasi 86%, menunjukkan tingkat kecocokan prediksi dengan label aktual yang cukup tinggi. Namun, variasi performa antar kelas mengindikasikan perlunya peningkatan, terutama pada klasifikasi sentimen positif dan netral.

4.6. Confusion Matrik

Evaluasi performa model klasifikasi secara mendalam dilakukan melalui analisis *confusion matrik*. Matrik ini menyajikan rekapitulasi hasil prediksi dengan membandingkan antara kelas aktual (*ground truth*) dan kelas prediksi yang dihasilkan oleh model.



Gambar 3. Confusion Matrik Naïve Bayes

Hasil evaluasi model klasifikasi sentimen pada isu "Indonesia Gelap" menunjukkan performa yang tidak seimbang antar kelas. Model ini menunjukkan akurasi tertinggi pada kelas negatif dengan berhasil mengklasifikasikan 426 data secara tepat. Sebaliknya, performa menurun drastis pada kelas netral dan positif, di mana masing-masing hanya 10 dan 4 data yang terklasifikasi dengan benar. Analisis kesalahan menunjukkan bahwa mayoritas data netral keliru diidentifikasi sebagai negatif. Temuan ini mengindikasikan bahwa model memiliki kecenderungan (bias) dan akurasi yang lebih tinggi dalam mendeteksi sentimen negatif dibandingkan sentimen lainnya.

4.7. Analisis Visualisasi

Visualisasi hasil merupakan salah satu tahap penting dalam analisis data, khususnya untuk memberikan gambaran yang lebih intuitif terhadap pola dan informasi yang terkandung dalam data isu “Indonesia Gelap”.

4.8. WordCloud

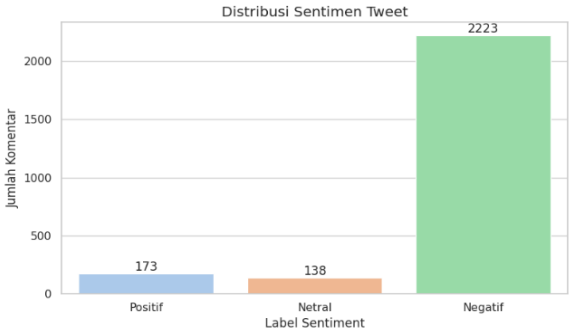
Visualisasi *WordCloud* memberikan wawasan kualitatif tentang kata-kata yang dominan pada setiap sentimen, bisa dilihat pada gambar 4 yang menyajikan kata-kata dominan pada isu “indonesia gelap”.

Gambar 4. Hasil WordCloud

Gambar 4 menunjukkan visualisasi WordCloud dari kata-kata yang paling sering terlihat bahwa kata-kata seperti "indonesia gelap", "rakyat", "negara", dan "darurat" muncul secara dominan, menunjukkan topik utama percakapan publik. Kata-kata seperti "tuntut", "perintah", dan "aksi" juga cukup mencolok, menunjukkan adanya protes, kritik, dan keinginan untuk perubahan. Secara keseluruhan, temuan WordCloud sebelumnya menunjukkan bahwa diskusi tentang masalah ini cenderung mengandung kritik sosial dan politik yang negatif.

4.9. Diagram Batang

Diagram batang digunakan untuk menunjukkan distribusi jumlah data di setiap kategori sentimen. Diagram ini menggambarkan jumlah tweet yang dikelompokkan ke dalam kategori negatif, netral, dan positif.

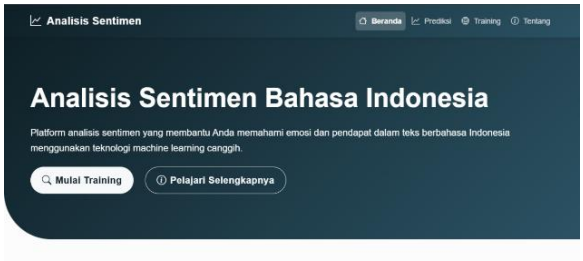


Gambar 5. Diagram Batang

Gambar 5 memperlihatkan distribusi sentimen dari seluruh tweet yang dianalisis, di mana sentimen negatif mendominasi secara signifikan dengan 2.223 tweet, dibandingkan dengan 173 tweet positif dan 138 tweet netral. Ketimpangan ini menunjukkan bahwa isu "Indonesia Gelap" cenderung memicu respons negatif dari pengguna, mencerminkan kekecewaan, kritik, atau ketidakpuasan publik terhadap kondisi yang diangkat.

4.10. UI (User Interface) Pengguna

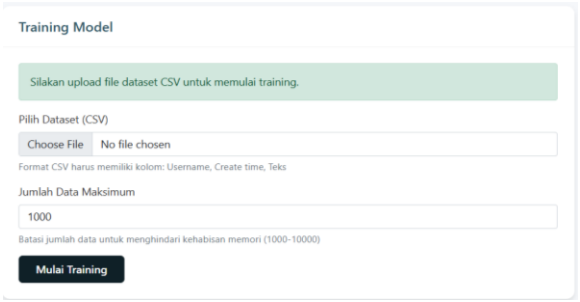
Antarmuka pengguna (UI) disediakan untuk menerapkan model analisis sentimen yang telah dibuat sebelumnya. UI ini membantu pengguna melakukan pelatihan model, melihat hasil analisis, dan membuat prediksi sentimen secara interaktif. Berikut adalah penjelasan untuk setiap halaman utama pada aplikasi.



Gambar 6. Halaman Utama

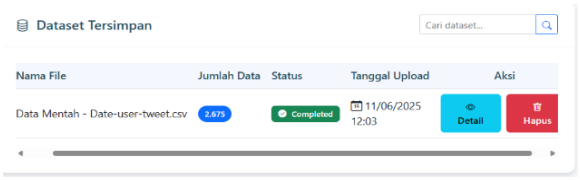
Gambar 6, menampilkan halaman utama (Beranda) aplikasi. Halaman ini berfungsi sebagai gerbang awal, memberikan deskripsi singkat mengenai platform analisis sentimen. Terdapat dua tombol utama: "Mulai Training" untuk pengguna yang ingin melatih model dengan dataset baru dan "Pelajari Selengkapnya" untuk informasi lebih lanjut. Navigasi

di bagian atas memungkinkan akses cepat ke menu "Beranda", "Prediksi", "Training", dan "Tentang".



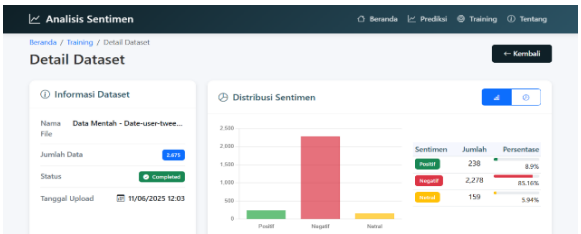
Gambar 7. Tampilan Halaman Training Model

Pada Gambar 7, di halaman ini pengguna dapat mengunggah dataset dalam format .csv melalui tombol "Choose File". File CSV harus memiliki kolom seperti Username, Create time, Teks. Supaya tidak terjadi error. Terdapat juga kolom untuk menentukan jumlah maksimum data yang akan diproses guna mengelola penggunaan memori. Setelah file dipilih dan jumlah data ditentukan, proses pelatihan dan analisis dapat dimulai dengan menekan tombol "Mulai Training".



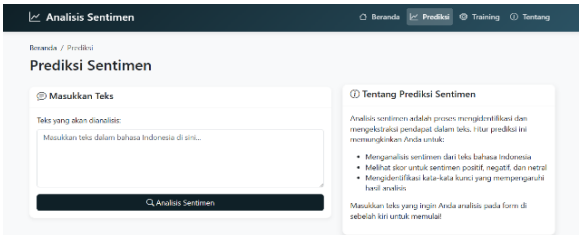
Gambar 8. Menu Dataset Tersimpan

Gambar 8 menunjukkan tampilan daftar dataset yang telah berhasil diunggah dan diproses. Setiap entri menampilkan informasi penting seperti nama file, jumlah data (jumlah data yang ditampilkan sudah melewati proses teks preprocessing), status proses (misalnya, "Completed"), dan tanggal unggah. Pengguna diberikan opsi untuk melihat detail hasil analisis melalui tombol "Detail" atau menghapus dataset dengan tombol "Hapus".



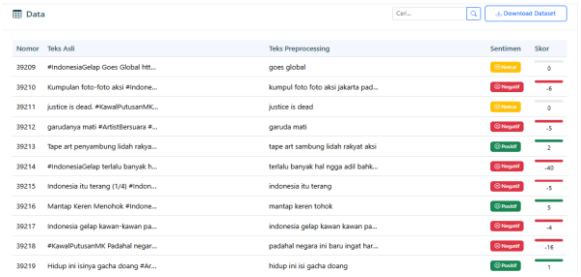
Gambar 9. Halaman Detail Dataset

Pada Gambar 9, tampialan halaman Detail Dataset yang dapat diakses setelah menekan tombol "Detail". Halaman ini merangkum informasi utama dari dataset yang dianalisis, termasuk nama file dan jumlah data. Di sisi kanan, terdapat visualisasi distribusi sentimen dalam bentuk diagram batang, lengkap dengan tabel persentase untuk sentimen Positif, Negatif, dan Netral.



Gambar 10. Halaman Prediksi

Aplikasi ini juga menyediakan fitur prediksi sentimen secara langsung yang ditampilkan pada Gambar 10. Pengguna dapat memasukkan kalimat atau teks baru berbahasa Indonesia ke dalam kotak teks yang tersedia. Dengan menekan tombol "Analisis Sentimen", sistem akan memproses masukan tersebut dan menampilkan hasil prediksinya.



Gambar 11. Halaman Tabel

Gambar 11. menunjukkan halaman detail yang menyajikan data dalam format tabel. Halaman ini menampilkan hasil lengkap dari proses analisis untuk setiap baris data, meliputi kolom "Teks Asli", "Teks Preprocessing", "Sentimen", dan "Skor". Fitur ini memungkinkan pengguna untuk memeriksa secara rinci bagaimana setiap *tweet* diproses dan diklasifikasikan oleh sistem, serta menyediakan fungsi pencarian dan unduh dataset.

5. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa algoritma Naïve Bayes dengan metode ekstraksi fitur CountVectorizer mampu mengklasifikasikan sentimen pada isu "Indonesia Gelap" dengan akurasi keseluruhan 86%. Hasil analisis menegaskan bahwa sentimen publik di media sosial X terhadap isu ini didominasi secara signifikan oleh opini negatif (87,7%), yang mengindikasikan adanya ketidakpuasan yang meluas. Model menunjukkan kinerja yang sangat baik untuk kelas mayoritas (negatif) tetapi kurang efektif untuk kelas minoritas (positif dan netral) karena ketidakseimbangan data. Untuk penelitian selanjutnya, disarankan untuk menerapkan teknik penyeimbangan data seperti SMOTE (*Synthetic Minority Over-sampling Technique*) untuk meningkatkan kinerja pada kelas minoritas. Selain itu, eksplorasi algoritma lain seperti SVM atau model berbasis *deep learning* (LSTM, BERT) serta metode deteksi sarkasme yang lebih canggih dapat menjadi arah pengembangan untuk menghasilkan analisis yang lebih akurat dan komprehensif.

DAFTAR PUSTAKA

[1] D. Rahmawati and T. Hidayat, "Analisis Kepuasan Mahasiswa Terhadap Sistem Informasi Akademik (Sina) Universitas Islam Syekh Yusuf Tangerang Berdasarkan Instrumen End User Computing Satisfaction (Eucs)," *Jurnal Teknik Informatika Unis)*, vol. 9, no. 1, pp. 2252–5351, Apr. 2021, doi: doi.org/10.33592/jutis.v9i1.842.

[2] D. Irenniza, A. Putri, F. T. Saputra, and R. Hardiyanti, "Pengaruh Penggunaan Media Sosial Twitter Terhadap Pemenuhan Kebutuhan Informasi (Survei Terhadap Pengikut Akun @Habisnontonfilm)," *Jurnal Ilmiah Wahana Pendidikan*, vol. 10, no. 8, pp. 410–418, Mar. 2024, doi: 10.5281/zenodo.11107309.

[3] S. Benammar, "'SO FUCKING CUTE IM DYING': The use of hyperbole on X," *Etudes de stylistique anglaise*, vol. 19, 2024, doi: 10.4000/11rfp.

[4] K. Hermansson, "The symbolic potential of security: on collective emotions in Swedish criminal policy discourse," *Emotions and Society*, vol. 4, no. 3, pp. 395–410, Nov. 2022, doi: 10.1332/263169021X16517125618359.

[5] M. A. Al Montaser *et al.*, "Sentiment analysis of social media data: Business insights and consumer behavior trends in the USA," *Edelweiss Applied Science and Technology*, vol. 9, no. 1, pp. 515–535, Jan. 2025, doi: 10.55214/25768484.v9i1.4164.

[6] Wilonotomo, B. Mulyawan, M. Ryanindityo, M. A. Syahrin, and F. Y. Triana, "Text Mining Algorithm Naive Bayes Classifier to Improve Quality Sentiment Analysis Passport Mobile Application," *Journal of Social and Political Sciences*, vol. 7, no. 1, Mar. 2024, doi: 10.31014/aior.1991.07.01.475.

[7] A. Yadav and D. V. Rao, "Fake News Detection Using Naive Bayes Classifier: A Comparative Study," *Journal of Management and Service Science (JMSS) A2Z Journals" Journal of Management and Service Science*, vol. 03, no. 022, pp. 1–14, Apr. 2023, doi: 10.54060/jmss.2023.22.

[8] E. Darwisah Harahap and R. Kurniawan, "Analisis Sentimen Komentar Terhadap Kebijakan Pemerintah Mengenai Tabungan Perumahan Rakyat (TAPERA) Pada Aplikasi X Menggunakan Metode Naïve Bayes," *Jurnal Teknik Informatika UNIKA Santo Thomas (JTIUST)*, pp. 2657–1501, Jun. 2024, doi: 10.54367/jtiust.v9i1.

[9] J. Khab Sulaiman Dalam, A. Oktavia Praneswara, N. Cahyono, and U. Amikom Yogyakarta, "Analisis Sentimen Ulasan Aplikasi TikTok Shop Seller Center di Google Playstore Menggunakan Algoritma Naive Bayes," *Indonesian Journal of Computer Science Attribution*, vol. 12, no. 6, pp. 3925–3940, Dec. 2023, doi: 10.33022/ijcs.v12i6.3473.

[10] M. S. Bahri and H. Kuswanto, "ANALISIS SENTIMEN PENGGUNA MEDIA SOSIAL X TERHADAP KONFLIK BERKEPANJANGAN PALESTINA DENGAN ISRAEL," *JIKA (Jurnal Informatika)*, vol. 8, no. 3, p. 261, Jul. 2024, doi: 10.31000/jika.v8i3.10632.

[11] Y. Akbar *et al.*, "Analisa Sentimen Pada Media Sosial"X" Pencarian Keyword ChatGPT Menggunakan Algoritma K-Nearest Neighbors (KNN)," *Jurnal Indonesia : Manajemen Informatika dan Komunikasi (JIMIK)*, vol. 5, no. 3, Sep. 2024, doi: 10.35870/jimik.v5i3.1016.

[12] R. Azhar, A. Surahman, and C. Juliane, "Analisis Sentimen Terhadap Cryptocurrency Berbasis Python TextBlob Menggunakan Algoritma Naïve Bayes," *Jurnal Sains Komputer & Informatika (J-SAKTI*, vol. 6, no. 1, pp. 267–281, Mar. 2022, doi: 10.30645/j-sakti.v6i1.443.

[13] D. Prastyo, D. Irawan, and I. H. Mursyidin, "Klasifikasi Sentimen Komentar YouTube dengan NLP pada Debat Pilkada Banten 2024," *bit-Tech (Binary Digital -Technology)*, vol. 7, no. 2, pp. 413–421, Dec. 2024, doi: 10.32877/bt.v7i2.1833.

[14] A. Z. Amrullah, A. Sofyan Anas, M. Adrian, and J. Hidayat, "Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square," *Jurnal BITe*, vol. 2, no. 1, pp. 40–44, Jul. 2020, doi: 10.30812/bite.v2i1.804.

[15] N. Hasdyna, "PREDICTIVE MODELING OF BROILER CHICKEN PRODUCTION USING THE NAIVE BAYES CLASSIFICATION ALGORITHM," *Jurnal Techno Nusa Mandiri*, vol. 21, no. 1, pp. 22–28, Mar. 2024, doi: 10.33480/techno.v21i1.5354.

[16] A. M. Alnasrawi, A. M. N. Alzubaidi, and A. A. Al-Moadhen, "Improving sentiment analysis using text network features within different machine learning algorithms," *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 1, pp. 405–412, Feb. 2024, doi: 10.11591/eei.v13i1.5576.

[17] Y. Bilgen and M. Kaya, "EGMA: Ensemble Learning-Based Hybrid Model Approach for Spam Detection," *Applied Sciences (Switzerland)*, vol. 14, no. 21, Nov. 2024, doi: 10.3390/app14219669.

[18] M. I. Fikri, T. S. Sabrila, Y. Azhar, and U. M. Malang, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *SMATIKA Jurnal*, vol. 10, no. 2, pp. 71–76, Dec. 2020, doi: 10.32664/smatika.v10i02.455