Analasis Sentimen Opini Masyarkat Pada Komentar Youtube Terhadap Kereta Cepat Jakarta – Bandung Menggunakan Metode Naive Bayes

1st Zidan Fathannul Khoiri
Sistem Informasi, Fakultas Rekayasa Industri
Universitas Telkom Purwokerto
Purwokerto, Indonesia
zidanfathannul@student.telkomuniversity.ac.id

2nd Sena Wijayanto, S.Pd., M.T
Sistem Informasi, Fakultas Rekayasa Industri
Universitas Telkom Purwokerto
Purwokerto, Indonesia
senawijayanto@telkomuniversity.ac.id

***Abstrak* — kereta Cepat Jakarta-Bandung (Whoosh) merupakan proyek transportasi terbesar di Indonesia yang memicu berbagai tanggapan dari masyarakat, baik dukungan maupun kritik. Opini publik terhadap proyek ini banyak ditemukan di media sosial, khususnya pada kolom komentar *YouTube*. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap proyek tersebut menggunakan algoritma *Multinomial Naive Bayes*. Data diperoleh dari komentar pada beberapa video YouTube yang membahas Kereta Cepat Jakarta-Bandung berjumlah 4161, kemudian melalui tahapan prapemrosesan, pelabelan sentimen, dan klasifikasi. Sentimen dikategorikan menjadi dua, yaitu positif dan negatif. Dari total 4.058 komentar yang dianalisis, sebanyak 3.000 komentar (75,49%) termasuk sentimen positif dan 974 komentar (24,50%) tergolong negatif. Komentar positif umumnya mengungkapkan rasa bangga dan optimisme terhadap kemajuan infrastruktur nasional, sedangkan komentar negatif cenderung menyoroti aspek pembiayaan, hutang, dan dampaknya terhadap perekonomian negara. Hasil evaluasi menunjukkan bahwa model mempertahankan performa yang baik pada data uji dengan akurasi sebesar 81%. Penelitian ini memberikan gambaran umum tentang persepsi publik serta dapat dijadikan acuan dalam menilai opini masyarakat dan perencanaan proyek infrastruktur di masa mendatang.**

***Kata kunci*— analisis sentimen, *youtube*, kereta cepat jakarta-bandung, *naive bayes classifier*.**

I. PENDAHULUAN

Kemajuan teknologi telah mendorong transformasi signifikan di berbagai sektor, termasuk transportasi. Di Indonesia, kereta api menjadi moda transportasi yang semakin diminati karena mampu mengangkut banyak penumpang dan terbebas dari kemacetan. Salah satu inovasi terbaru adalah proyek Kereta Cepat Jakarta–Bandung (KCJB) atau Whoosh, yang dianggap sebagai simbol kemajuan teknologi dan efisiensi dalam mobilitas masyarakat [1]

Proyek KCJB memberikan manfaat seperti pemangkasan wa ktu tempuh dan peningkatan konektivitas antarwilayah, namun juga menimbulkan kontroversi, terutama terkait pembengkakan biaya dan dampak sosial terhadap masyarakat sekitar [2]. Hal ini memunculkan beragam opini publik yang banyak disampaikan melalui media sosial, khususnya YouTube. YouTube menjadi saluran efektif bagi masyarakat untuk menyuarakan pendapat melalui komentar. Komentar-komentar ini mengandung informasi berharga yang dapat dianalisis untuk memahami sentimen publik terhadap proyek KCJB. Analisis sentimen merupakan metode untuk mengidentifikasi opini positif, negatif, atau netral dalam teks, dan sering digunakan dalam riset media sosial[3]. Salah satu metode yang banyak digunakan dalam analisis sentimen adalah Naive Bayes Classifier (NBC), karena sederhana, cepat, dan memiliki tingkat akurasi yang baik[4]. Berdasarkan latar belakang tersebut, penelitian ini mengangkat judul: "Analisis Sentimen Opini Publik pada Komentar YouTube tentang Kereta Cepat Jakarta–Bandung Menggunakan Metode Naive Bayes Classifier." Penelitian ini bertujuan untuk mengetahui persepsi publik terhadap proyek KCJB dengan memanfaatkan komentar-komentar di YouTube sebagai data utama.

II. KAJIAN TEORI

A. Analisis Sentimen

Analisis sentimen (*sentiment analysis* atau *opinion mining*) adalah studi komputasional yang bertujuan untuk mengenali, mengekspresikan, dan mengevaluasi opini, sentimen, sikap, emosi, subjektivitas, maupun pandangan yang terdapat dalam suatu teks. Analisis ini memungkinkan identifikasi sentimen positif dan negatif terhadap individu, institusi, produk, maupun isu tertentu. Hasil analisis biasanya dikelompokkan ke dalam dua kategori utama: positif dan

negatif. Selain itu, analisis sentimen juga dapat diperluas untuk mengenali pihak atau kelompok yang menjadi sumber utama dari sentimen tersebut, baik yang bersifat positif maupun negatif.[5]

B. Youtube

Sebagai platform digital, *Youtube* menyediakan berbagai fitur interaktif, seperti menonton, memberikan reaksi, membagikan, serta mengomentari video secara gratis. Pengguna dapat mengunggah video yang membahas berbagai topik, mulai dari hiburan, edukasi, ulasan produk, hingga isu-isu sosial dan kebijakan publik. Seiring meningkatnya jumlah video yang dipublikasikan, kolom komentar menjadi wadah utama bagi pengguna untuk menyampaikan opini dan sentimen mereka terhadap topik yang dibahas dalam video. Komentar-komentar ini mencerminkan opini publik yang beragam, baik relevan maupun tidak, dan menjadi sumber data penting untuk menganalisis pandangan masyarakat terhadap suatu isu [6].

C. Text Mining

Text mining merupakan proses eksplorasi dan analisis data teks tidak terstruktur dengan tujuan menemukan pola, konsep, kata kunci, dan informasi penting lainnya. Metode ini dikenal pula sebagai *knowledge discovery in textual databases (KDT)*, yaitu proses penggalian pengetahuan baru dari kumpulan teks dalam skala besar (*corpus*) yang kemudian diolah menjadi informasi yang bermakna dan aplikatif [7][8].

D. Smote

SMOTE adalah teknik *resampling* yang digunakan untuk mengatasi ketidakseimbangan kelas dalam dataset dengan cara menambahkan sampel sintetis pada kelas minoritas. Berbeda dengan *undersampling*, SMOTE tidak mengurangi data dari kelas mayoritas sehingga tidak terjadi kehilangan informasi. Teknik ini efektif dalam meningkatkan kinerja model klasifikasi, terutama dalam mengenali pola dari kelas yang jarang muncul [9].

E. Naive Bayes

Naive Bayes adalah metode klasifikasi statistik berbasis probabilitas yang memprediksi kemungkinan suatu data termasuk dalam kelas tertentu. Metode ini didasarkan pada *Teorema Bayes* dengan asumsi bahwa setiap fitur dalam data bersifat independen satu sama lain. Asumsi ini membuat perhitungan menjadi lebih sederhana dan efisien, bahkan ketika jumlah data pelatihan terbatas[10] Salah satu varian populer dari algoritma ini adalah *Multinomial Naive Bayes*, yang banyak digunakan dalam pengolahan bahasa alami (*Natural Language Processing*). Algoritma ini mempertimbangkan frekuensi kemunculan kata dalam dokumen untuk melakukan klasifikasi. Model ini memperhitungkan tidak hanya keberadaan kata, tetapi juga seberapa sering kata tersebut muncul dalam teks, sehingga cocok untuk analisis sentimen berbasis teks.[11] Multinomial Naive Bayes dapat dirumuskan dengan cara berikut gabungkan

$$\boldsymbol{P(c) = \frac{N_c}{N_{Doc}}} \quad \textbf{(1)}$$

Keterangan Probabilitas *Posterior*:

C : Kategori atau kelas.

Doc : Dokumen.

N_c : Banyaknya kategori C pada dokumen latih.

N_{doc} : Banyaknya keseluruhan dokumen latih yang digunakan.

$$\boldsymbol{P(x_i|C_k) = \frac{Jumlah\ Kata\ x_i\ di\ kelas\ C_k + \alpha}{Total\ semua\ kata\ di\ kelas\ C_k + (\alpha.V)}} \quad \textbf{(2)}$$

Keterangan kata dalam kelas :

Jumlah kata x_i : Berapa kali kata tersebut muncul di kelas tersebut.

Total Semua kata : Jumlah total kata dalam kelas tersebut.

α : Angka tambahan (biasanya 1) untuk mencegah pembagian nol

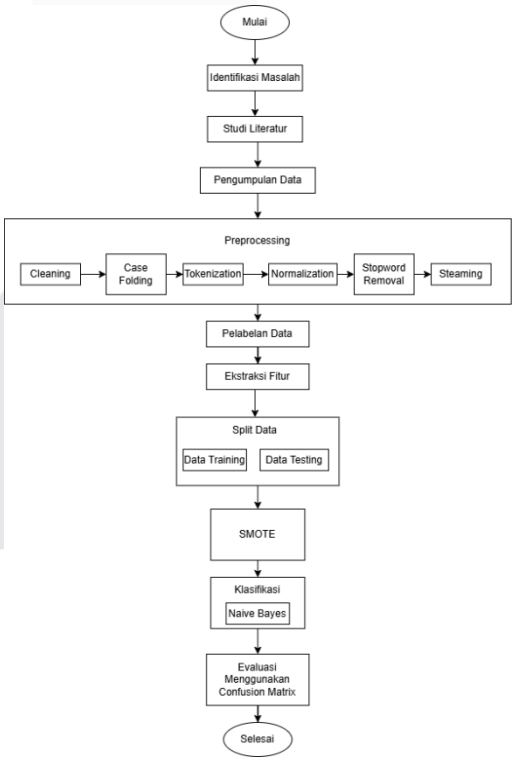
V : Jumlah total kata unik di semua kelas (*Vocabulary*)

$$\boldsymbol{P(c,d) = p(c)\Pi\ 1 \le i \le count(v,d)p(w_i,c)} \quad \textbf{(3)}$$

Keterangan kata dalam kelas :

$count\ (w_{i,}c)$: jumlah kata unik pada dokumen.

III. METODE



GAMBAR 1
(DIAGRAM ALIR PENELITIAN)

Penelitian ini diawali dengan identifikasi dan perumusan masalah, penyusunan latar belakang, penetapan tujuan, serta studi literatur terkait analisis sentimen menggunakan

algoritma Machine Learning, khususnya *Naive Bayes*. Literatur diperoleh dari *Google Scholar* dengan batasan publikasi lima tahun terakhir. Data dikumpulkan dari komentar-komentar pada video berita tentang proyek Kereta Cepat Jakarta–Bandung di YouTube menggunakan *Youtube* API melalui *Google Colab* dan *Python*, menghasilkan 4161 baris data dalam format *CSV*. Tahap berikutnya adalah *preprocessing* data, yang meliputi pembersihan teks (*cleaning*), pengubahan huruf menjadi kecil (*case folding*), pemisahan kata (*tokenization*), normalisasi kata tidak baku (*normalization*), penghapusan kata umum (*stopword removal*), serta pengembalian kata ke bentuk dasar (*stemming*). Setelah dibersihkan, data diberi label sentimen positif atau negatif menggunakan kamus SenticNet berbasis Bahasa Indonesia berdasarkan nilai polaritas kata. Proses *feature extraction* dilakukan dengan metode *TF-IDF (Term Frequency-Inverse Document Frequency)*, yang menghitung seberapa sering kata muncul dalam dokumen dan seberapa jarang kata tersebut muncul di seluruh dokumen, sehingga menghasilkan bobot yang berguna dalam klasifikasi. Data kemudian dibagi menjadi data latih dan data uji dengan rasio 80:20 untuk menghindari overfitting atau underfitting. Mengingat data tidak seimbang, diterapkan teknik *Synthetic Minority Oversampling Technique* (*SMOTE*) untuk menambah sampel sintetis pada kelas minoritas agar distribusi data lebih seimbang. Selanjutnya, algoritma *Multinomial Naive Bayes* digunakan untuk membangun model klasifikasi, karena cocok untuk data berbasis teks dan mempertimbangkan frekuensi kemunculan kata. Model ini digunakan untuk memprediksi apakah suatu komentar bersifat positif atau negatif. Evaluasi model dilakukan dengan *confusion matrix*, yang membandingkan hasil prediksi dengan data aktual dan menghasilkan metrik evaluasi seperti akurasi, presisi, *recall,* dan *F1-score* guna menilai performa klasifikasi komentar terhadap proyek Kereta Cepat Jakarta–Bandung.

IV. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan melalui platform *Youtube* dengan memanfaatkan *Youtube Data API v3*. *API* ini memungkinkan peneliti untuk mengakses metadata video, termasuk komentar-komentar yang ditulis oleh pengguna pada video tertentu, data yang berhasil terkumpul sebanyak 4161.

	publishedAt	authorDisplayName	textDisplay	likeCount
0	2023-10-27T20:19:24Z	@deddysondjafri1870	Sy kerja di bandung, seminggu sekali pulang ke...	0
1	2023-09-13T17:40:01Z	@guylesmana	Harus jujur utk mengatakan bahwa trayek KCIC a...	1
2	2023-08-15T07:48:56Z	@olehsoleh8097	PROYEK LEGEG GAYA GAYA AN TIDAK ADA UNTUNGNYA	0
3	2023-08-13T06:21:34Z	@anastiaththariq7689	Harus beroperasi berapa abad ya untuk bisa mei...	0
4	2023-08-13T01:22:59Z	@hendriceria7827	DiPerancis ada kereta Maglev... Akhirnya ki...	0
...	...	...	...	...
202	2023-09-16T09:10:17Z	@holilisholehi4699	Silahkan, anda mau jalan kakipun monggo gk ada...	4
203	2023-09-16T06:30:26Z	@silverrayleigh5356	Aku mah bebas meninggal tiap ada kerusuhan...<...	0
204	2023-09-16T06:29:39Z	@monaaza5159	Jakarta Pati ada ngk ya????	2
205	2023-09-16T07:52:18Z	@dyon7630	Desa belahan kadrun ya eta, bye jelata	0
206	2023-09-16T08:38:41Z	@monaaza5159	@@dyon7630 uuuhhhh	0

207 rows × 4 columns

GAMBAR 2 :
(HASIL DATA)

B. Preprocessing

Data yang sudah dikumpulkan akan menjalani tahap preprocessing. Tujuan dari proses preprocessing ini adalah untuk membersihkan dan mempersiapkan data tersebut agar siap digunakan dalam analisis lebih lanjut. Berikut adalah penjelasan rinci mengenai proses tersebut:

a) Cleaning

untuk menghapus elemen-elemen yang tidak diperlukan dalam analisis. Proses ini mencakup penghilangan karakter khusus, emoji, simbol non-alfabet, angka, tanda baca, serta spasi yang berlebihan.

TABEL 1
(CLEANING)

Sebelum	Sesudah
Walaupun hutang, tpi tk mudah mngmbil kbjkan, yg resikonya mngkrak dan hujatan bahkan makian yg ekstrem dan kejam, bravo jokowiâ ¤â ¤â ¤	Walaupun hutang, tpi tk mudah mngmbil kbjkan, yg resikonya mngkrak dan hujatan bahkan makian yg ekstrem dan kejam, bravo jokowi

b) Case Folding

Pada tahap case folding, semua huruf dalam teks diubah menjadi huruf kecil. Proses ini bertujuan untuk menyamakan bentuk kata, sehingga perbedaan dalam penggunaan huruf kapital tidak memengaruhi arti kata tersebut.

TABEL 2
(CASE FOLDING)

Sebelum	Sesudah
Walaupun hutang, tpi tk mudah mngmbil kbjkan, yg resikonya mngkrak dan hujatan bahkan makian yg ekstrem dan kejam, bravo jokowi	walaupun hutang tpi tk mudah mngmbil kbjkan yg resikonya mngkrak dan hujatan bahkan makian yg ekstrem dan kejam bravo jokowi

c) Normalization

Pada tahap *normalization*, teks yang telah melalui proses *case folding* disesuaikan lagi dengan mengubah berbagai bentuk kata tidak baku menjadi bentuk kata baku. Normalisasi dilakukan untuk menyamakan variasi penulisan kata, seperti singkatan, bahasa gaul, atau kesalahan ketik yang umum terjadi pada teks komentar media sosial.

TABEL III : Normalization

Sebelum	Sesudah
walaupun hutang tpi tk mudah mngmbil kbjkan	Walaupun hutang tapi tidak mudah mengambil kebijakan

Sebelum	Sesudah
yg resikonya mngkrak dan hujatan bahkan makian yg ekstrem dan kejam bravo jokowi	yang resikonya mangkrak dan hujatan bahkan makian yang ekstrem dan kejam semangat jokowi

d) Tokenization

Pada tahap tokenisasi, teks yang telah dinormalisasi dipecah menjadi potongan-potongan kecil yang disebut token. Setiap token mewakili kata, simbol, atau tanda baca, sehingga memudahkan proses analisis selanjutnya.

TABEL 4
(TOKENIZATION)

Sebelum	Sesudah
walaupun hutang tapi tidak mudah mengambil kebijakan yang resikonya mangkrak dan hujatan bahkan makian yang ekstrem dan kejam semangat jokowi	['walaupun', 'hutang', 'tapi', 'tidak', 'mudah', 'mengambil', 'kebijakan', 'yang', 'resikonya', 'mangkrak', 'dan', 'hujatan', 'bahkan', 'makian', 'yang', 'ekstrem', 'dan', 'kejam', 'semangat', 'jokowi']

e) Stopword Removal

Dalam proses stopword removal, kita menghapus kata-kata umum yang sering muncul dalam teks tetapi tidak memberikan makna yang signifikan, seperti kata sambung, preposisi, atau sapaan. Langkah ini bertujuan untuk mengurangi gangguan (noise) dalam data dan memfokuskan analisis pada kata-kata yang lebih relevan.

TABEL 5
(STOPWORD REMOVAL)

Sebelum	Sesudah
['walaupun', 'hutang', 'tapi', 'tidak', 'mudah', 'mengambil', 'kebijakan', 'yang', 'resikonya', 'mangkrak', 'dan', 'hujatan', 'bahkan', 'makian', 'yang', 'ekstrem', 'dan', 'kejam', 'semangat', 'jokowi']	walaupun hutang tidak mudah mengambil kebijakan resikonya mangkrak hujatan bahkan makian ekstrem kejam semangat jokowi

f) Stemming

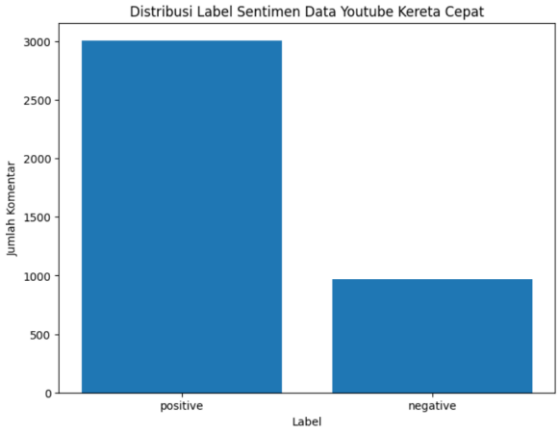
Dalam tahap stemming, kata-kata dalam teks dipulihkan ke bentuk dasar dengan cara menghapus imbuhan seperti awalan, sisipan, dan akhiran. Tujuan dari proses ini adalah untuk menstandarkan variasi kata sehingga analisis dapat dilakukan dengan lebih konsisten dan tepat.

TABEL 6
(STEMMING)

Sebelum	Sesudah
walaupun hutang tidak mudah mengambil kebijakan resikonya mangkrak hujatan bahkan makian ekstrem kejam semangat jokowi	walaupun hutang tidak mudah ambil bijak resiko mangkrak hujat bahkan maki ekstrem kejam semangat jokowi

C. Pelabelan

Tahap berikutnya adalah pelabelan. Dalam penelitian ini, pelabelan data dilakukan secara otomatis dengan menggunakan skor polaritas dari kamus *SenticNet ID*. Setiap komentar akan dikategorikan sebagai sentimen positif atau negatif, yang kemudian digunakan sebagai target dalam pelatihan model analisis sentimen.



GAMBAR 3
(HASIL PELABELAN)

Berdasarkan hasil pelabelan sentimen menggunakan pendekatan leksikon, analisis komentar YouTube terkait Kereta Cepat Jakarta–Bandung menunjukkan bahwa mayoritas bersentimen positif, yaitu sebanyak 3.000 komentar (75,49%), sedangkan 974 komentar (24,50%) tergolong negatif. Temuan ini mengindikasikan bahwa secara umum masyarakat memberikan dukungan terhadap proyek tersebut, meskipun masih ada komentar yang bersifat kritis.

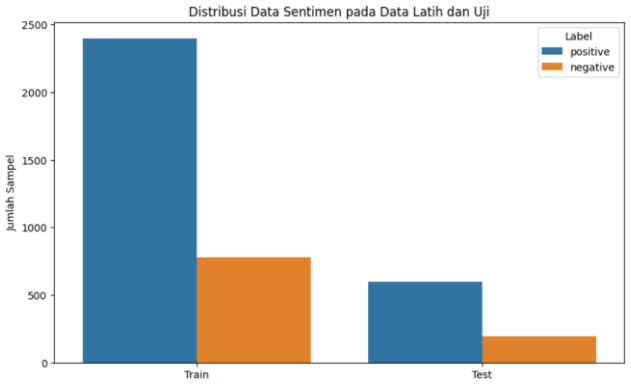
D. Ekstraksi Fitur

Proses transformasi data teks menjadi bentuk numerik dilakukan menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency) dengan bantuan *TfidfVectorizer*. Pengaturan yang digunakan meliputi max_df=0.85 untuk mengabaikan kata yang terlalu sering muncul, min_df=3 untuk menyaring kata yang muncul di setidaknya tiga dokumen, max_features=4500 untuk membatasi jumlah fitur unik, dan ngram_range=(1,1) untuk menggunakan unigram. Setelah dilatih pada data teks dari kolom *cleaned_senticnet*, setiap komentar diubah menjadi vektor angka dan disimpan dalam variabel x_tf_idf, yang kemudian digunakan sebagai input untuk model klasifikasi sentimen.

E. Split Data

Proses pemisahan data dilakukan dengan membagi data fitur hasil TF-IDF ke dalam dua bagian, yaitu 80% sebagai data latih dan 20% sebagai data uji, menggunakan fungsi train_test_split() dengan parameter random_state=42 untuk menjaga konsistensi hasil. Variabel x menyimpan fitur, sedangkan y menyimpan label sentimen. Setelah pembagian, ditampilkan jumlah serta distribusi kelas pada data latih guna memastikan pembagian data telah seimbang dan

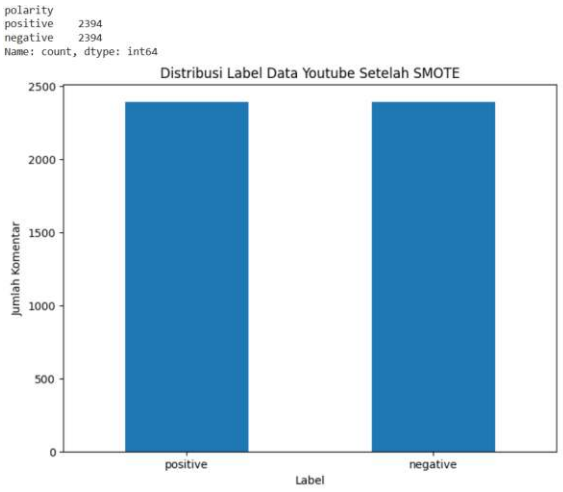
representatif. Tahap ini penting agar model dapat belajar secara optimal dan diuji secara adil.[12]



GAMBAR 4
(HASIL SPLIT DATA)

F. Balancing

Setelah proses penyeimbangan data menggunakan teknik *SMOTE,* jumlah data latih untuk label positif dan negatif menjadi seimbang, masing-masing sebanyak 2.394 komentar, dengan total 4.788 data. Teknik ini berhasil menambah data sintetis pada kelas minoritas (negatif), sehingga model memiliki peluang yang sama dalam mempelajari kedua kelas sentimen dan menghasilkan klasifikasi yang lebih adil dan akurat.



GAMBAR 5
(HASIL *SMOTE)*

G. Pengujian Model

Model Multinomial Naive Bayes digunakan untuk mengklasifikasikan komentar YouTube ke dalam dua kelas sentimen, yaitu positif dan negatif. Algoritma ini dipilih karena cocok untuk data teks yang telah dikonversi ke bentuk numerik menggunakan TF-IDF dan efektif dalam menangani klasifikasi berbasis frekuensi kata. Model dilatih pada data latih yang telah diseimbangkan dengan teknik SMOTE, kemudian digunakan untuk memprediksi sentimen pada data uji. Evaluasi performa dilakukan menggunakan metrik akurasi, presisi, recall, dan f1-score. Berdasarkan hasil evaluasi, model menunjukkan akurasi sekitar 81% pada data uji, yang menandakan kemampuannya dalam mengenali pola sentimen secara umum. Namun, model cenderung lebih akurat dalam mengklasifikasikan sentimen positif dibandingkan negatif.

Model Performance on Testing Data (MultinomialNB)
Accuracy: 0.8075471698113208

	precision	recall	f1-score	support
Negative	0.57	0.78	0.66	189
Positive	0.92	0.82	0.87	606
accuracy			0.81	795
macro avg	0.75	0.80	0.76	795
weighted avg	0.84	0.81	0.82	795

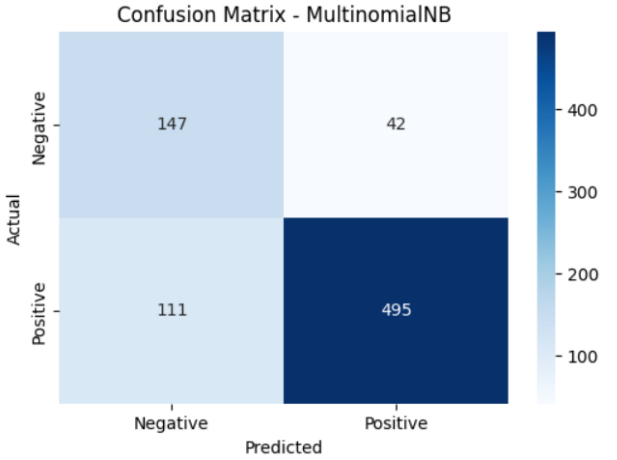
GAMBAR 6
(HASIL PENGUJIAN MODEL)

H. Evaluasi Hasil Pengujian

Hasil prediksi model pada data uji dibandingkan dengan label sebenarnya menggunakan confusion matrix untuk mengevaluasi akurasi klasifikasi. Berdasarkan hasil, model berhasil mengklasifikasikan 495 komentar positif dan 147 komentar negatif secara tepat, namun masih terjadi kesalahan klasifikasi pada 111 komentar positif yang diprediksi sebagai negatif dan 42 komentar negatif yang diprediksi sebagai positif. Visualisasi confusion matrix menunjukkan performa model yang cukup baik, terutama dalam mengenali komentar positif. Model Multinomial Naive Bayes mencapai akurasi sebesar 81%, dengan kinerja yang lebih tinggi pada sentimen positif (precision 0,92, recall 0,82, F1-score 0,87) dibandingkan sentimen negatif (precision 0,57, recall 0,78, F1-score 0,66). Hal ini mengindikasikan bahwa meskipun model efektif mengenali komentar positif, kemampuannya dalam mendeteksi komentar negatif masih perlu ditingkatkan.

V. KESIMPULAN

Implementasi algoritma Multinomial Naive Bayes dalam analisis sentimen terhadap komentar YouTube mengenai proyek Kereta Cepat Jakarta–Bandung berhasil dilakukan dengan akurasi mencapai 81% pada data pengujian, menunjukkan efektivitas metode ini dalam mengklasifikasikan opini publik berbasis teks. Hasil pelabelan menunjukkan bahwa mayoritas komentar, yaitu 3.000 komentar (75,49%), bersentimen positif yang umumnya mencerminkan dukungan terhadap kemajuan infrastruktur dan kebanggaan nasional, sementara 974 komentar (24,50%) tergolong negatif. Sentimen negatif tersebut umumnya berkaitan dengan kekhawatiran terhadap beban utang negara serta dampak sosial seperti relokasi warga. Temuan ini mengindikasikan bahwa meskipun proyek mendapat respons positif secara umum, masih terdapat kelompok masyarakat yang menyuarakan kritik terhadap risiko-risiko tertentu dari proyek tersebut.

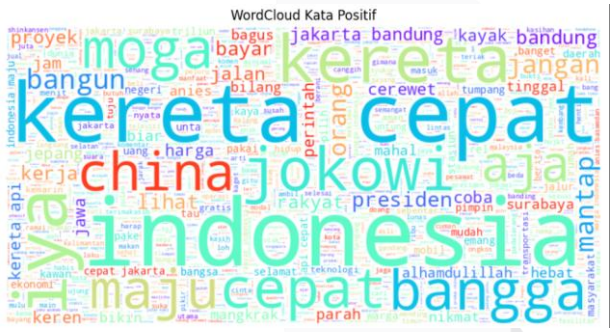


Gambar 7 : Hasil Evaluasi Model

I. Implikasi Hasil Analisis Sentimen

1) Wordcloud Sentimen Positif

Berdasarkan analisis terhadap 4.058 komentar YouTube tentang proyek Kereta Cepat Jakarta–Bandung, ditemukan bahwa 75,60% bersentimen positif dan 24,39% bersentimen negatif. Mayoritas komentar positif menyoroti peningkatan konektivitas dan efisiensi transportasi, sementara komentar negatif mencerminkan kekhawatiran terkait biaya, dampak lingkungan, dan efektivitas proyek, menunjukkan adanya pandangan yang beragam di masyarakat.



Gambar 8 : Hasil Wordcloud Positif

2) Wordcloud Sentimen Negatif

Visualisasi *wordcloud* sentimen positif terhadap proyek Kereta Cepat Jakarta–Bandung menampilkan kata-kata dominan seperti *kereta*, *cepat*, *bangga*, *mantap*, *maju*, *Indonesia*, dan *jokowi*, yang mencerminkan antusiasme, dukungan, serta harapan masyarakat terhadap kemajuan infrastruktur nasional. Kata-kata seperti *moga* dan *terus* menunjukkan optimisme publik agar proyek ini berhasil dan membawa manfaat jangka panjang, sebagaimana tercermin dalam komentar positif seperti "kalau bukan pak Jokowi tidak ada kereta cepat" dan "maju terus Indonesia".



Gambar 9 : Hasil Wordcloud Negatif

REFERENSI

[1] J. L. Rizky and W. Gata, "Analisis Sentimen Media Sosial Youtube Kereta Cepat (Whoosh) Menggunakan Algoritma Bidirectional-LSTM," 2024.

[2] R. N. Afifah and Y. R. Bangun, "Exploring Public Perception Towards Jakarta-Bandung High Speed Rail," Scitepress, Feb. 2020, pp. 413–422.

[3] J. Supriyanto, D. Alita, and A. R. Isnain, "Penerapan Algoritma K-Nearest Neighbor (K-NN) Untuk Analisis Sentimen Publik Terhadap Pembelajaran Daring," *Jurnal Informatika dan Rekayasa Perangkat Lunak*, vol. 4, no. 1, pp. 74–80, Mar. 2023.

[4] A. R. Fitriansyah, "Analisis Sentimen Terhadap Pembangunan Kereta Cepat Jakarta-Bandung Pada Media Sosial Twitter Menggunakan Metode SVM dan GloVe Word Embedding," 2023.

[5] Z. N. Aulia, G. Kuncoro Jati, and I. Santoso, "Analisis Sentimen TanggapanPublic Mengenai E - Tilang Melalui Media Sosial Youtube Menggunakan Algoritma Naive Bayes," 2023.

[6] E. F. Saputra and M. R. Pribadi, "Analisis Sentimen Komentar Pada Kanal Youtube The Lazy Monday Menggunakan Algoritma Naive Bayes," 2023.

[7] F. Fathonah and A. Herliana, "Penerapan Text Mining Analisis Sentimen Mengenai Vaksin Covid - 19 Menggunakan Metode Naïve Bayes," *Jurnal Sains dan Informatika*, vol. 7, no. 2, pp. 155–164, Dec. 2021.

[8] A. Firdaus and W. I. Firdaus, "Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan)," 2021.

[9] K. Pramayasa, I. Md, D. Maysanjaya, G. Ayu, and A. Diatri Indradewi, "Analisis Sentimen Program Mbkm Pada Media Sosial Twitter Menggunakan KNN Dan SMOTE," 2023.

[10] H. Herlawati, R. T. Handayanto, P. D. Atika, F. N. Khasanah, A. Y. P. Yusuf, and D. Y. Septia, "Analisis Sentimen Pada Situs Google Review dengan Naïve Bayes dan Support Vector Machine," *Jurnal Komtika (Komputasi dan Informatika)*, vol. 5, no. 2, pp. 153–163, Nov. 2021.

[11] Yuyun, Nurul Hidayah, and Supriadi Sahibu, "Algoritma Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Pemerintah Terhadap

Penanganan Covid-19 Menggunakan Data Twitter," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 4, pp. 820–826, Aug. 2021.

[12] A. Wijayanto *et al.*, "ANALISIS KLASIFIKASI KEPUASAN PENUMPANG MASKAPAI PENERBANGAN MENGGUNAKAN ALGORITMA NAÏVE BAYES." [Online]. Available: www.kaggle.com