

# Perbandingan Algoritma *Naïve Bayes* Dan *K-Nearest Neighbor* (KNN) Pada Video Youtube Mengenai *Global Warming*

Aldi Himawan<sup>1</sup>, Castaka Agus Sugianto<sup>2</sup>

<sup>1,2</sup>Politeknik TEDC Bandung, Jl. Politeknik-Pesantren KM2 Cimahi Utara – Cimahi – Jawa Barat - Indonesia  
e-mail: aldihimawan00@gmail.com.<sup>1</sup>,Castaka@poltektedc.ac.id<sup>2</sup>

## INFORMASI ARTIKEL

### Sejarah Artikel:

Diterima Redaksi : 19 Juli 2024

Revisi Akhir : 04 November 2024

Diterbitkan Online : 30 November 2024

### Kata Kunci:

Pemanasan Global, *YouTube*, Analisis Sentimen, *Naïve Bayes*, *K-Nearest Neighbor* (KNN), Sentimen Komentar.

### Korespondensi :

Telepon / Hp : 081282076508

E-mail : [aldihimawan00@gmail.com](mailto:aldihimawan00@gmail.com)

## A B S T R A K

Pemanasan global merupakan salah satu tantangan lingkungan utama yang disebabkan oleh peningkatan emisi gas rumah kaca. Media sosial, khususnya *YouTube*, menjadi platform penting dalam menyebarkan informasi dan menciptakan kesadaran tentang isu-isu ini. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap video *YouTube* yang membahas pemanasan global, khususnya video yang dibuat oleh Najwa Shihab. Dua metode klasifikasi digunakan, yaitu *Naïve Bayes* dan *K-Nearest Neighbor* (KNN), untuk mengklasifikasikan komentar pengguna menjadi kategori positif atau negatif. Data komentar dikumpulkan menggunakan teknik web scraping pada Februari 2024, dan dianalisis menggunakan metode SEMMA Data Mining Process. Hasil analisis menunjukkan bahwa metode *Naïve Bayes* dan KNN memberikan kinerja yang kompetitif dalam mengklasifikasikan sentimen komentar-komentar ini. Evaluasi dilakukan dengan membagi dataset komentar menjadi beberapa rasio pelatihan dan pengujian, Evaluasi dilakukan dengan membagi dataset komentar menjadi beberapa rasio pelatihan dan pengujian: 70:30, 80:20, dan 90:10. Hasil analisis menggunakan rasio 90:10 yang sudah dilakukan menggunakan *10-fold cross validation* menunjukkan bahwa algoritma *Naïve Bayes* memperlihatkan kinerja yang kompetitif dengan akurasi sebesar 79.17%, presisi sebesar 80.76% *recall* sebesar 79.17%, dan *f1-score* sebesar 78.68%. Sementara itu, *K-Nearest Neighbor* juga menunjukkan kinerja yang kompetitif, dengan akurasi sebesar 76.06%, presisi sebesar 76.70%, *recall* sebesar 76.06%, dan *f1-score* sebesar 75.98%. Hasil pengujian *T-Test* menunjukan perbedaan yang signifikan antara algoritma *Naïve Bayes* dan *K-Nearest Neighbor* dengan nilai probabilitas 0.0001 kurang dari nilai alpha yaitu 0.05 yang menunjukan bahwa algoritma *Naïve Bayes* lebih unggul performanya. Selain itu, analisis sentimen menunjukkan bahwa mayoritas komentar memiliki sentimen negatif, mencerminkan bahwa video tersebut berhasil meningkatkan kesadaran masyarakat mengenai isu pemanasan global dan memicu diskusi yang konstruktif.

## 1. PENDAHULUAN

Pemanasan global adalah suatu kondisi ketidakseimbangan ekosistem bumi yang disebabkan oleh kenaikan suhu rata-rata di atmosfer, laut, dan daratan. Dalam seratus tahun terakhir, suhu rata-rata permukaan bumi telah naik sebesar  $0.74 \pm 0.18$  °C. Kenaikan ini disebabkan oleh peningkatan emisi gas rumah kaca, seperti karbon dioksida, metana, dinitrogen oksida, hidrofluorokarbon, perfluorokarbon, dan sulfur heksafluorida di atmosfer. Emisi-emisi ini terutama dihasilkan dari pembakaran bahan bakar fosil (minyak bumi dan batu bara) serta aktivitas penggundulan dan pembakaran hutan [1].

Pemanasan global adalah salah satu isu lingkungan utama yang dihadapi dunia saat ini. Pemanasan global terkait dengan peningkatan suhu rata-rata permukaan bumi. Kenaikan suhu ini disebabkan oleh radiasi sinar matahari yang mencapai atmosfer bumi, kemudian sebagian dari radiasi tersebut berubah menjadi energi panas dalam bentuk sinar inframerah yang diserap oleh udara dan permukaan bumi [1]. Saat ini, isu pemanasan global banyak dibahas di media sosial, termasuk di *YouTube*.

*YouTube* merupakan salah satu media sosial yang sangat populer di kalangan masyarakat Indonesia. Popularitas *YouTube* diperkirakan akan terus meningkat seiring dengan bertambahnya jumlah pengguna. Pada pertengahan tahun 2017, *YouTube* mencatat jumlah penonton terdaftar sebanyak 1,5 miliar pengguna bulanan. Pada tahun 2019, lembaga riset pasar Statista memprediksi bahwa jumlah pengguna *YouTube* akan mencapai 1,8 miliar pada tahun 2021 [2]. Sebuah riset menyebutkan bahwa 88% pengguna di Indonesia menganggap *YouTube* sebagai sumber utama mereka untuk mencari informasi melalui konten video [3]. Banyaknya pengguna aktif *YouTube* di Indonesia membuka peluang besar bagi para content creator Indonesia untuk memanfaatkan platform ini sebagai media hiburan, sumber informasi, hingga untuk berbagi ide-ide tertentu. Salah satu content creator terkenal di *YouTube* Indonesia adalah Najwa Shihab.

Najwa Shihab adalah seorang presenter, jurnalis, dan content creator di Indonesia yang membahas berbagai isu, termasuk politik, hukum, sosial, religi, dan isu-isu terkini. Kanal *YouTube*-nya didirikan sejak tahun 2018, dengan ide awal berasal dari Najwa Shihab sendiri,

yang juga merupakan Co-Founder Narasi. Setelah empat bulan, kanal YouTube Najwa Shihab berhasil mendapatkan 250 ribu pengikut dan memperoleh silver button. Saat ini, jumlah pengikut kanal YouTube-nya telah mencapai 9,51 juta. Dalam lima tahun, Najwa Shihab telah berhasil menarik jutaan penonton dari seluruh konten yang dipublikasikannya. Salah satu video yang populer adalah "2050, Andai Kita Tidak Melakukan Apa Apa". Video ini menjadi topik hangat saat dirilis, dengan lebih dari 684 ribu kali penayangan dan 2.004 komentar. Video tersebut membahas dampak pemanasan global di masa depan dan menerima banyak komentar positif, netral, hingga negatif.

Kombinasi antara dampak pemanasan global di masa depan dengan video yang dibuat oleh Najwa Shihab tentang dampak pemanasan global sangat menarik perhatian publik, terutama masyarakat Indonesia. Banyaknya komentar dari publik mencerminkan berbagai tanggapan, dan tanggapan ini dapat dianalisis menggunakan teknik analisis sentimen. Analisis sentimen, atau opinion mining, adalah proses mengolah, mengekstrak, dan memahami data tekstual secara otomatis untuk mendapatkan informasi sentimen dari suatu kalimat atau opini [4].

Penelitian mengenai analisa sentimen telah dilakukan oleh beberapa peneliti seperti oleh Samsir et al pada tahun 2021. Penelitian tersebut berjudul "Analisis Sentimen Pembelajaran Daring Pada Twitter dimasa Pandemi COVID-19 Menggunakan Metode *Naïve Bayes*". Penelitian tersebut mengenai respon dari masyarakat mengenai perubahan mendadak penerapan pembelajaran daring mengingat adanya pandemi Covid-19 [5]. Kemudian terdapat penelitian analisa sentimen dari Castaka dan Shandy pada tahun 2022 yang berjudul "A Covid-19 Sentiment Analysis on Twitter Using *K-Nearest Neighbors*". Penelitian tersebut menjelaskan mengenai analisis sentimen masyarakat Indonesia terhadap topik Covid-19 [6].

Dengan penelitian penelitian sebelumnya yang banyak menggunakan algoritma seperti metode *Naïve Bayes* dan *K-Nearest Neighbor* dalam melakukan analisis sentimen. Algoritma Naive Bayes merupakan metode pengkategorian menggunakan probabilitas dan statistik [7]. Naive Bayes classifier adalah metode statistik untuk klasifikasi yang dapat memperkirakan probabilitas keanggotaan dalam suatu kelas. Metode ini terbukti memiliki tingkat akurasi yang tinggi dan dapat bekerja dengan cepat ketika diterapkan pada dataset yang besar [8]. Sedangkan algoritma *K-Nearest Neighbor* adalah metode klasifikasi yang menggunakan label kategori dari dokumen pelatihan, seperti dokumen uji, dsb [9]. Pada dasarnya *K-Nearest Neighbor* dapat mengklasifikasikan item berdasarkan data pelatihan yang lebih dekat [10]. Peneliti dalam penelitian ini tertarik untuk menggunakan metode *Naïve Bayes* dan *K-Nearest Neighbor*. Dari referensi-referensi penelitian tersebut, penelitian ini akan membandingkan metode *Naïve Bayes* dan *K-Nearest Neighbor* dalam menganalisis sentimen masyarakat.

Dengan alasan alasan diatas, penelitian ini akan dilakukan untuk menganalisis opini-opini masyarakat terhadap video Najwa Shihab mengenai *global warming* pada *platform* Youtube dengan menggunakan algoritma *Naïve Bayes* dan *K-Nearest Neighbor*.

## 2. METODE PENELITIAN

### 2.1. Menentukan Masalah

Masalah penelitian adalah inti dari sebuah penelitian, menggambarkan kondisi yang dapat diamati dengan jelas dan akurat. Masalah ini kemudian menimbulkan pertanyaan yang perlu dijawab, yang mengarahkan penulis pada kegiatan penelitian. Masalah yang diangkat dalam penelitian berasal dari konsep realitas yang perlu diselesaikan. Kekuatan sebuah isu terletak pada keunggulannya, bukan pada anggapan apakah masalah itu ada atau tidak. Identifikasi masalah yang penulis sampaikan adalah langkah awal dari penelitian. Tahap penentuan masalah telah dijelaskan dalam latar belakang yang membahas sentimen masyarakat terhadap pemanasan global menggunakan algoritma *Naïve Bayes* dan *K-Nearest Neighbor* yang berasal dari permasalahan yang ada.

### 2.2. Menentukan Tujuan dan Ruang Lingkup

Menetapkan tujuan dan ruang lingkup penelitian melibatkan penentuan apa yang ingin dicapai dalam penelitian ini serta batasan-batasan yang diperlukan agar penelitian tidak terlalu luas dan lebih efektif.

### 2.3. Mencari Litelature

Penelitian literatur atau referensi dilakukan untuk mencari berbagai sumber yang berkaitan dengan Analisis Sentimen, *Text Mining*, Klasifikasi, *Naïve Bayes*, dan *K-Nearest Neighbor*. Metode yang digunakan adalah studi literatur. Dalam penelitian ini, metode penelitian dokumenter digunakan untuk mengumpulkan dan meneliti data dari buku, internet, dan jurnal yang relevan dengan penelitian yang sedang dilakukan.

Data yang dikumpulkan oleh penulis dari metode penelitian dokumenter adalah :

1. Analisis Sentimen
2. *Youtube*
3. Pemanasan Global
4. *Text Mining*
5. Klasifikasi
6. Metode SEMMA
7. *Cross Validation*
8. *Lexicon Based*
9. *Web Scraping*
10. *Python*

### 2.4. Pengumpulan Data

Tahap yang paling penting dalam penelitian adalah pengumpulan data. Pengumpulan data adalah proses mengumpulkan data untuk keperluan penelitian, sehingga peneliti tidak akan mendapatkan hasil jika data tidak dikumpulkan. Proses pengumpulan data menjadi krusial dalam penelitian karena metode yang digunakan untuk mengumpulkan data akan menentukan kualitas

dan keakuratan data yang diperoleh. Dalam proses penelitian ini, penulis akan menggunakan metode pengumpulan data berikut:

1) Data Primer

Data primer merupakan data yang diperoleh langsung dari sumber data yang relevan dengan penelitian yang dilakukan. Pada penelitian ini data yang digunakan diperoleh langsung dari media sosial yaitu salah satu video Youtube Najwa Shihab.

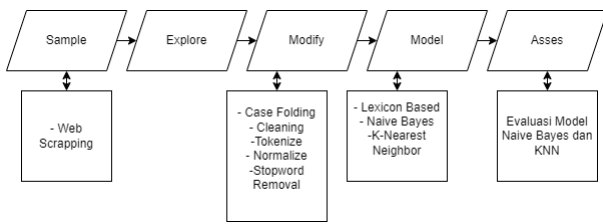
2) Data Sekunder

Data sekunder adalah data yang didapat secara tidak langsung bersumber dari dokumentasi, literatur, buku, jurnal dan informasi lainnya yang berhubungan dengan rumusan masalah dan kebutuhan penulisan.

2.5. Metode Yang Digunakan

Penelitian ini mengacu kepada metode SEMMA Data Mining Process untuk metodologi penelitian. Fokus didalam metode SEMMA ini adalah modifikasi, pemodelan, dan penambangan data [7].

Metode SEMMA terdiri dari sample, explore, modify, model and asses seperti pada Gambar 1.



Gambar 1. Proses Pengolahan Data

3. HASIL DAN PEMBAHASAN

3.1. Sample

Pada tahap ini dilakukan pengumpulan dataset yang bersumber dari komentar video YouTube Najwa Shihab yang berjudul “2025, Andai Kita Tidak Melakukan Apa-Apa” dengan teknik scraping komentar yang dilakukan pada Februari 2024. Komentar yang didapat berjumlah 2.004 komentar dengan menggunakan App Script yang memanfaatkan fungsi YouTube API komentar-komentar tersebut disimpan didalam file.csv.

Langkah-langkah yang akan dilakukan oleh peneliti termasuk login atau masuk ke akun Google untuk mengakses fitur-fitur yang disediakan oleh Google, seperti YouTube API, app script, dan aplikasi pendukung lainnya. Selanjutnya, peneliti dapat memanfaatkan fitur-fitur tersebut untuk melakukan proses crawling data menggunakan bahasa pemrograman Python. Tabel 1 menunjukkan dataset yang diambil dari komentar YouTube.

Tabel 1. Dataset dari komentar youtube

No	User	Komentar
1.	@yuyuwahyudinofficial7132	Kalau lihat ini mengapa ibu kota negara harus pindah? Padahal dengan pindah

No	User	Komentar
		akan menghilangkan hutan dan ekologi. entahlah!
2.	@musicnostalgic99	siapaun yang baca komen saya, tanamlah pohon, tolong...
.	.	.
.	.	.
.	.	.
2.004	@MariaF.Anmuni-ww9ku	bumi makin Tua Pemanasan Global dam efec rumah kaca Wabah penyakit dan perubahan iklim

3.2. Explore

Pada tahap ini, peneliti mendeskripsikan data yang diperoleh dari hasil crawling atau scraping komentar video YouTube dari kanal Najwa Shihab yang berjudul "2050, Andai Kita Tidak Melakukan Apa-Apa". Data yang didapat kemudian divisualisasikan dalam bentuk grafik berdasarkan waktu untuk menampilkan informasi secara visual.

3.3. Modify

Pada tahap Modify di penelitian ini adalah tahapan untuk persiapan data atau pre-processing. Tahapan yang akan dilakukan oleh peneliti terdiri dari case folding, cleaning, tokenize, dan normalize.

1) Case Folding

Langkah pertama pada tahap modifikasi dalam penelitian ini adalah case folding. Tahap case folding dilakukan untuk mengubah semua karakter dalam dataset menjadi huruf kecil. Proses case folding dilakukan menggunakan Python. Setelah proses ini, dataset yang dihasilkan tidak lagi mengandung huruf kapital. Tabel 2 berikut ini menunjukkan hasil dari case folding.

Tabel 2. Hasil Case Folding

SebelumCaseFolding	SesudahCaseFolding
Kalau lihat ini mengapa ibu kota negara harus pindah? padahal dengan pindah akan menghilangkan hutan dan ekologi. entahlah!	kalau lihat ini mengapa ibu kota negara harus pindah? padahal dengan pindah akan menghilangkan hutan dan ekologi. entahlah!
bumi makin Tua Pemanasan Global dam efec rumah kaca Wabah penyakit dan perubahan iklim	bumi makin tua pemanasan global dam efec rumah kaca wabah penyakit dan perubahan iklim

2) *Cleaning*

Langkah selanjutnya adalah membersihkan data dari komponen yang tidak relevan dan tidak memiliki makna seperti karakter ASCII, angka, link, hashtag, URL, tanda baca, dan whitespace. Tabel 3 berikut ini menunjukkan hasil pembersihan pada dataset dalam penelitian ini.

Tabel 3. Hasil *Cleaning*

Sebelum <i>Cleaning</i>	Sesudah <i>Cleaning</i>
Terimakasih mba nana, kontennya sangat menginspirasi, semoga bisa bermanfaat untuk kita semuaâ	terimakasih mba nana, kontennya sangat menginspirasi, semoga bisa bermanfaat untuk kita semua

3) *Tokenize*

Tahap selanjutnya adalah tokenisasi, yaitu proses memecah kalimat dari komentar *YouTube* atau dataset menjadi potongan kata atau token. Ini dilakukan untuk mengetahui kemunculan kata-kata tersebut. Tabel 4 berikut ini menunjukkan hasil *tokenisasi* pada dataset dalam penelitian ini.

Tabel 4. Hasil *Tokenize*

Sebelum <i>Tokenize</i>	Sebelum <i>Tokenize</i>
siapapun yang baca komen saya tanamlah pohon tolong	['siapapun', 'yang', 'baca', 'komen', 'saya', 'tanamlah', 'pohon', 'tolong']

4) *Normalize*

Tahap berikutnya adalah normalisasi. Dalam penelitian ini, normalisasi dilakukan untuk menstandarkan kata-kata yang memiliki makna yang sama dengan mengubah penulisan kata yang disingkat atau tidak baku. Penelitian ini menggunakan kamus normalisasi yang diperoleh dari kamus NLP (Neuro Linguistic Programming) Bahasa Indonesia Resource [11]. Tabel 5 berikut ini menunjukkan hasil normalisasi pada dataset dalam penelitian ini.

Tabel 5. Hasil *Normalize*

Sebelum <i>Tokenize</i>	Sesudah <i>Tokenize</i>
['siapapun', 'yang', 'baca', 'komen', 'saya', 'tanamlah', 'pohon', 'tolong']	['siapa', 'pun', 'yang', 'baca', 'komentar', 'saya', 'tanam', 'lah', 'pohon', 'tolong']

5) *Stopword Removal*

Tujuan dari tahap *Stopword Removal* adalah untuk menghapus kata-kata umum yang sering digunakan namun tidak memberikan pengaruh signifikan terhadap sentimen suatu kalimat. Proses penghapusan *stopword* dalam penelitian ini menggunakan library dari Sastrawi yang menyertakan *corpus stopwords* bahasa Indonesia. Tabel 6 berikut menampilkan hasil dari tahap *Stopword Removal* dalam dataset penelitian ini.

Tabel 6. Hasil *Stopword Removal*

Sebelum <i>Stopword Removal</i>	Sesudah <i>Stopword Removal</i>
['siapa', 'pun', 'yang', 'baca', 'komentar', 'saya', 'tanam', 'lah', 'pohon', 'tolong']	['siapa', 'yang', 'baca', 'komentar', 'saya', 'tanam', 'pohon', 'tolong']

3.4. *Model*

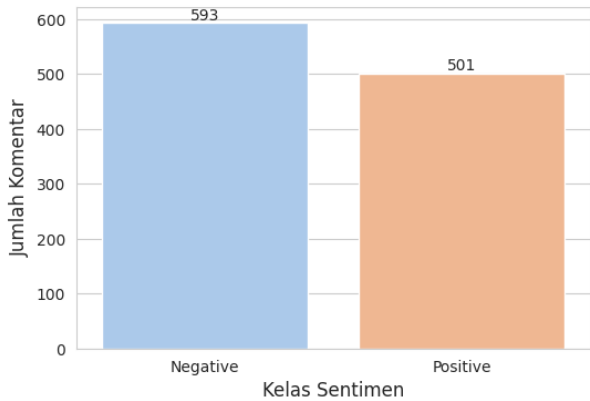
Dalam tahap *Model* didalam penelitian ini melakukan pelabelan dengan metode lexicon based untuk mendapatkan label sentimen didalam dataset dan menggunakan 2 pemodelan klasifikasi dengan algortima *Naive Bayes* dan *K-Nearest neighbor*.

1) *Lexicon Based*

Metode berbasis lexicon yang digunakan dalam penelitian ini bertujuan untuk melakukan klasifikasi kelas pada dataset. Setiap komentar atau dokumen dalam dataset diklasifikasikan menjadi dua kelas, yaitu positif atau negatif. Kata-kata dalam komentar yang terdapat dalam kamus *lexicon* akan diberi nilai skor. Jumlah skor pada suatu komentar menentukan apakah labelnya positif atau negatif. Tabel 7 berikut menunjukkan hasil dari dataset yang telah melalui pelabelan dengan metode berbasis *lexicon*. Gambar 2 merupakan jumlah dataset komentar yang sudah diklasifikasi menggunakan metode *Lexicon Based*.

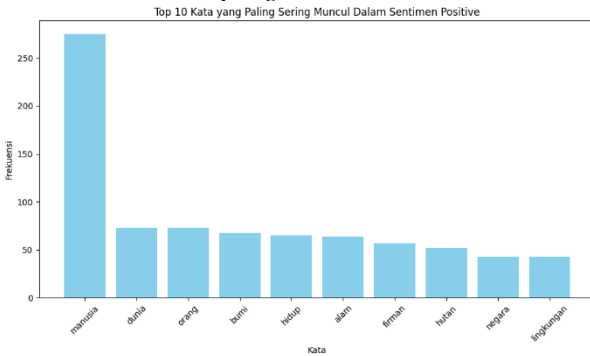
Tabel 7. Hasil Metode Lexicon Based

No	Komentar	Label
1	kalau lihat ini mengapa ibu kota negara harus pindah padahal dengan pindah akan menghilangkan hutan dan ekologi. Entahlah	Negatif
2	siapapun yang baca komen saya, tanamlah pohon, tolong...	Positif
3	sejak dulu kita semua tinggal di planet bumi jika planet ini rusak kita tidak tau lagi mau tinggal di mana.	Negatif
4	terimakasih mba nana, kontennya sangat menginspirasi, semoga bisa bermanfaat untuk kita semua	Positif
5	dimana logikanya bu, awal bilang air banjir merajalela setinggi lutut di semua wilayah, tapi ko bisa kekurangan air logikamu gak nyambung bu, seperti baju yg kamu pake kwkwkw	Negatif
.	.	.
.	.	.
.	.	.
1784	bumi makin tua pemanasan global dam efek rumah kaca wabah penyakit dan perubahan iklim	Negatif

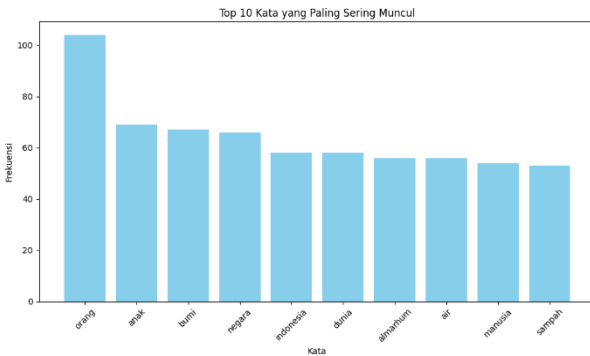


Gambar 2. Jumlah Komentar Setelah Dilakukan *Lexicon Based*

Setelah data sudah dilakukan pelabelan, Pada tahap ini, Dataset ini kemudian di visualisasikan dalam bentuk *Word Freq Positive* dan *Word Freq Negative*. Gambar 3 Berikut menampilkan hasil *Word Freq Positive* dan Gambar 4 *Word Freq Negative*.

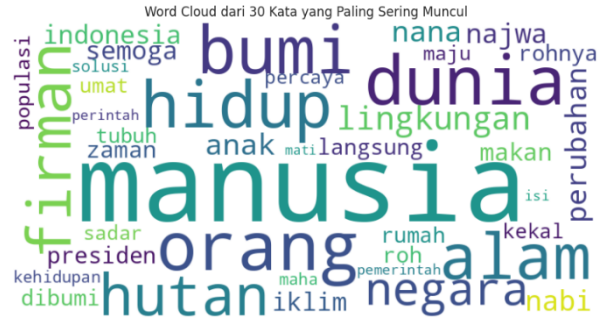


Gambar 3. *Word Freq Positive*



Gambar 4. *Word Freq Negative*

Selanjutnya data di visualisasikan dalam bentuk *Word Cloud* sesuai masing-masing kelas *positive* dan *negative*. Gambar 5 Berikut menampilkan hasil *Word Cloud Positive* dan Gambar 6 *Word Cloud Negative*.



Gambar 5. *Word Cloud Positive*



Gambar 6. *Word Cloud Negative*

### 2) *Naive Bayes*

Pada tahap ini peneliti akan membagi dataset menjadi data latih (*training*) dan data uji (*test*). Tabel 8 Merupakan rasio perbandingan data latih dan data uji yang akan dilakukan menggunakan algoritma *Naive Bayes*.

Tabel 8. Perbandingan Data Latih dan Data Uji *Naive Bayes*

Data Latih	Data Uji
70%	30%
80%	20%
90%	10%

### 3) *K-Nearest Neighbor*

Pada tahap ini peneliti akan membagi dataset menjadi data latih (*training*) dan data uji (*test*). Tabel 9 Merupakan rasio perbandingan data latih dan data uji yang akan dilakukan menggunakan algoritma *K-Nearest Neighbor*.

Tabel 9. Perbandingan Data Latih dan Data Uji *K-Nearest Neighbor*

Data Latih	Data Uji
70%	30%
80%	20%
90%	10%

### 3.5. Asses

Tahap terakhir yaitu *Asses* merupakan tahap untuk dilakukannya evaluasi dari setiap model pada penelitian ini. Hasil evaluasi berisikan nilai-nilai akurasi, presisi, *recall*, *f1-score*, *confussion matrix* dan *10-fold cross validation*.

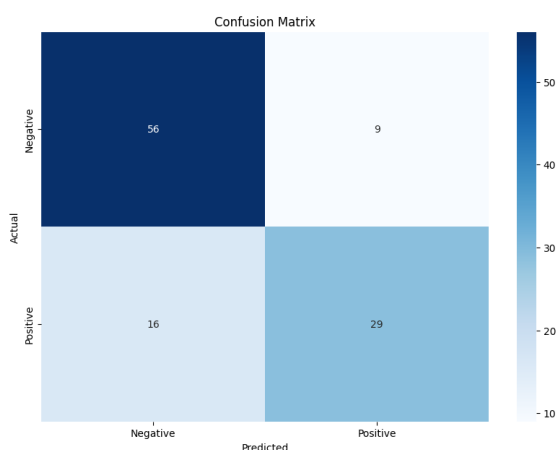
1) Naive Bayes

Pada metode *Naive Bayes* didalam penelitian ini, seperti yang sudah dijelaskan dalam tahap model, penelitian ini membagi 3 untuk data latih dan data uji dan kemudian diolah dnegan metode *Naive Bayes*. Tabel 10 merupakan hasil dari pengujian algoritma *Naive Bayes*.

Tabel 10. Hasil Akurasi, Presisi, Recall, F1-Score *Naive Bayes*

DL : DU	Akurasi	Presisi	Recall	F1-score
70:30	76.29%	78.33%	76.29%	75.32%
80:20	75.34%	76.04%	75.34%	74.53%
90:10	77.27%	77.18%	77.27%	76.89%

Kemudian untuk hasil dari *Confussion Matrix* dengan algoritma *Naive Bayes* pada rasio 90%:10% didapat bahwa prediksi benar sentimen positif (*True Positive*) ada sebanyak 29 data dan prediksi benar untuk sentimen negatif (*True Negative*) sebanyak 56 data. Gambar 7 merupakan hasil *Confussion Matrix Naive Bayes*.



Gambar 7. *Confussion Matrix Naive Bayes*

Untuk lebih memastikan keakuratan dan konsistensi performa algoritma *Naive Bayes*, dilakukan pengujian menggunakan *10-fold cross validation*. Hasil yang didapat setelah melakukan pengujian *10-fold cross validation* menunjukkan akurasi sebesar 79.17% , presisi sebesar 80.76%, *recall* sebesar 79.17% dan *f1-score* sebesar 78.68%.

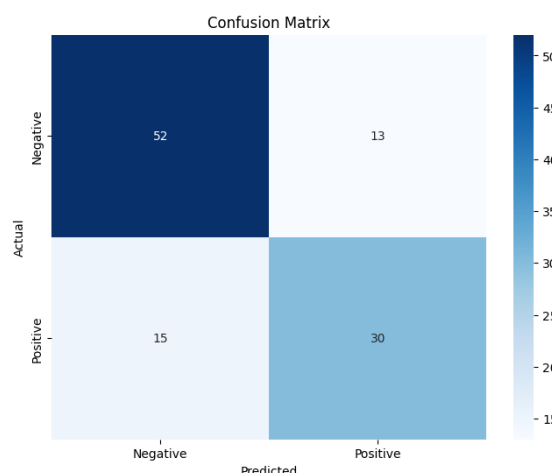
2) K-Nearest Neighbor

Pada metode *K-Nearest Neighbor* didalam penelitian ini, seperti yang sudah dijelaskan dalam tahap model, penelitian ini membagi 3 untuk data latih dan data uji dan kemudian diolah dnegan metode *K-Nearest Neighbor*. Tabel 11 merupakan hasil dari pengujian algoritma *Naive Bayes*.

Tabel 11. Hasil Akurasi, Presisi, Recall, F1-Score KNN

DL : DU	Akurasi	Presisi	Recall	F1-score
70:30	75.68%	75.63%	75.68%	75.63%
80:20	73.06%	72.89%	73.06%	72.85%
90:10	74.55%	74.55%	74.55%	74.55%

Kemudian untuk hasil dari *Confussion Matrix* dengan algoritma *K-Nearest Neighbor* pada rasio 90%:10% didapat bahwa prediksi benar sentimen positif (*True Positive*) ada sebanyak 30 data dan prediksi benar untuk sentimen negatif (*True Negative*) sebanyak 52 data. Gambar 8 merupakan hasil *Confussion Matrix K-Nearest Neighbor*.



Gambar 8. *Confussion Matrix K-Nearest Neighbor*

Untuk lebih memastikan keakuratan dan konsistensi performa algoritma *K-Nearest Neighbor*, dilakukan pengujian menggunakan *10-fold cross validation*. Hasil yang didapat setelah melakukan pengujian *10-fold cross validation* menunjukkan akurasi sebesar 76.06% , presisi sebesar 76.70%, *recall* sebesar 76.06% dan *f1-score* sebesar 75.98%.

3.6. Interpretasi Hasil

Data yang dikumpulkan sebanyak 2.004 komentar. Setelah terkumpul data melewati beberapa tahapan seperti *case folding*, *cleaning*, *tokenize*, *normalize*, *stopword removal*, dan pelabelan menggunakan *lexicon based*. Setelah melewati beberapa tahapan tersebut total data menjadi 1.094 diakrenakan dalam penelitian ini hanya mengambil data yang memiliki label positif dan negatif. Kemudian dilakukan pengujian menggunakan algoritma *Naive Bayes* dan *K-Nearest Neighbor* (KNN).

Pengujian dilakukan untuk mengukur performa masing-masing algoritma dalam mengklasifikasikan sentimen positif dan negatif dari komentar yang ada. Kinerja algoritma diukur menggunakan metrik akurasi, presisi, *recall*, dan *f1-score*. Tabel 12 merupakan hasil pengujian dari algoritma *Naive Bayes* dan *K-Nearest Neighbor* setelah dilakukan *10-fold cross validation*.

Tabel 12. Hasil *10-fold Cross Validation*

Algoritma	10-fold Cross Validation			
	Akurasi	Presisi	Recall	F1-Score
Naïve Bayes	79.17%	80.76%	79.17%	78.68%
K-Nearest Neighbor	76.06%	76.70%	76.06%	75.98%

Hasil dari pengujian *10-fold cross validation* menunjukkan bahwa algoritma *Naive Bayes* memiliki kinerja yang konsisten dengan akurasi rata-rata di atas 79%. Dan KNN memiliki kinerja yang cukup baik dengan akurasi rata-rata sekitar 76.06%. Metrik presisi, *recall*, dan *f1-score* juga menunjukkan hasil yang baik, menandakan bahwa kedua algoritma ini efektif dalam mengklasifikasikan sentimen pada komentar-komentar *YouTube* mengenai *global warming*.

Kemudian setelah dilakukan pengujian *T-Test* terhadap hasil *10-fold cross validation* dari kedua algoritma tersebut didapatkan hasil seperti pada Gambar 9. Hasil dari pengujian *T-Test* menunjukkan bahwa terdapat perbedaan yang signifikan antara algoritma *Naive Bayes* dan *K-Nearest Neighbor* dikarenakan nilai probabilitiknya 0.0001 kurang dari nilai alpha yaitu 0.05.

t-statistic : 4.811515866284426  
P-value : 0.00010614166791034315

Gambar 9. Hasil Pengujian *T-Test*

## 5 KESIMPULAN

Berdasarkan hasil pengujian yang telah dilakukan dalam penelitian ini, maka dapat disimpulkan bahwa :

1. Berdasarkan evaluasi, didapatkan bahwa kedua algoritma *Naive Bayes* menghasilkan hasil terbaik masing-masing nilai akurasi sebesar 79.17%, presisi sebesar 80.76%, *recall* sebesar 79.19%, dan *f1-score* sebesar 78.68%. Sedangkan pada algoritma *K-Nearest Neighbor* didapatkan nilai k terbaik adalah 11 dengan menghasilkan nilai akurasi sebesar 76.06%, presisi 76.70%, *recall* sebesar 76.06%, dan *f1-score* sebesar 75.98%. Dari hasil tersebut dapat disimpulkan bahwa kedua algoritma ini efektif dalam mengklasifikasikan sentimen pada komentar-komentar *Youtube* mengenai *global warming*.
2. Dari hasil pengujian *T-Test*, dapat disimpulkan bahwa algoritma *Naive Bayes* dan *K-Nearest Neighbor* memiliki perbedaan kinerja yang signifikan dalam konteks dataset dan pengujian yang telah dilakukan. Oleh karena itu, dalam penelitian ini algoritma *Naive Bayes* lebih unggul performanya dibandingkan dengan algoritma *K-Nearest Neighbor*.
3. Dari analisis yang dilakukan, ditemukan bahwa sentimen masyarakat terhadap video tersebut bervariasi. Sebagian besar komentar menunjukkan sentimen negatif dengan jumlah 593 komentar negatif dan 501 komentar positif, dengan pengguna yang merasa video tersebut informatif dan penting dalam meningkatkan kesadaran tentang dampak pemanasan global. Hasil ini menunjukkan bahwa video Najwa Shihab berhasil menarik perhatian dan memicu diskusi di kalangan masyarakat mengenai isu pemanasan global.

## DAFTAR PUSTAKA

- [1] H. Nik, T. Agus, and P. M. Wiji, "Strategi Penanggulangan Pemanasan Global Terhadap Dampak Laju Perekonomian Dalam Pandangan Islam," *J. Din. Ekon. Syariah*, vol. 9, no. 2, pp. 168–183, 2022.
- [2] M. Haryadi, "Pemanfaatan Youtube Sebagai Media Ajar Dalam Meningkatkan Minat Dan Motivasi Belajar," vol. 5, no. 1, pp. 135–159, 2019.
- [3] N. Laugu, "Eksistensi Media Sosial, Youtube, Instagram dan Whatsapp Ditengah Pandemi Covid-19 Dikalangan Masyarakat Virtual Indonesia," vol. 4, no. 1, pp. 41–57, 2020.
- [4] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," *INTEGER J. Inf. Technol.*, vol. 2, no. 1, pp. 32–41, 2017, doi: 10.31284/j.integer.2017.v2i1.95.
- [5] W. A. Prabowo and C. Wiguna, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naive Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 1, pp. 157–163, 2021, doi: 10.30865/mib.v5i1.2604.
- [6] C. A. Sugianto and S. Tresnawati, "A Covid-19 Sentiment Analysis on Twitter Using K-Nearest Neighbours," *J. Appl. Intell. Syst.*, vol. 7, no. 1, pp. 58–69, 2022, doi: 10.33633/jais.v7i1.5984.
- [7] D. Era, S. Andryana, and A. Rubhasy, "Perbandingan Algoritma Naive Bayes Dan K-Nearest Neighbor pada Analisis Sentimen Pembukaan Pariwisata Di Masa Pandemi Covid 19," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 7, no. 1, pp. 263–272, 2023.
- [8] T. H. Apandi and C. A. Sugianto, "Analisis Komparasi Machine Learning Pada Data Spam Sms," *J. TEDC*, vol. 12, no. 1, pp. 58–62, 2018.
- [9] M. Iqbal, A. Davy Wiranata, R. Suwito, and R. Faiz Ananda, "KLIK: Kajian Ilmiah Informatika dan Komputer Perbandingan Algoritma Naive Bayes, KNN, dan Decision Tree terhadap Ulasan Aplikasi Threads dan Twitter," *Media Online*, vol. 4, no. 3, pp. 1799–1807, 2023, doi: 10.30865/klik.v4i3.1402.
- [10] R. Puspita and A. Widodo, "Perbandingan Metode KNN, Decision Tree, dan Naive Bayes Terhadap Analisis Sentimen Pengguna Layanan BPJS," *J. Inform. Univ. Pamulang*, vol. 5, no. 4, pp. 646–654, 2021, doi: 10.32493/informatika.v5i4.7622.
- [11] W. Afifi, "Analisis sentimen pengguna twitter terhadap layanan internet pt indosat tbk dengan metode k-nearest neighbor (k-nn) dan naive bayes classifier (nbc)," *Repository.Uinjkt.Ac.Id*, 2022.