

ANALISIS SENTIMEN MASYARAKAT TERHADAP PELANTIKAN KABINET MERAH PUTIH PADA MEDIA SOSIAL MENGGUNAKAN ALGORITMA NAIVE BAYES

Ade Krisna S¹, Haycal saptahadi HS², Zidan Fathanul K³, Sukmadiningtyas⁴

¹²³⁴ Sistem Informasi, Universitas Telkom
Jl. D.I. Panjaitan No. 128 Purwokerto, 53147, Indonesia

¹adekrisna@student.telkomuniversity.ac.id
²haycalsaptahadi@student.telkomuniversity.ac.id
³zidanfathannul@student.telkomuniversity.ac.id
⁴sukmadiningtyasvxn@telkomuniversity.ac.id

Dikirim pada 22-11-2024, Direvisi pada 27-11-2024, Diterima pada 04-12-2024

Abstrak

Pelantikan Kabinet Merah Putih 2024 dibawah kepemimpinan Presiden Prabowo Subianto dan Wakil Presiden Gibran Rakabuming mendapat sorotan publik, memunculkan berbagai opini di media sosial. Penelitian ini bertujuan menganalisis sentimen opini publik mengenai pelantikan tersebut dengan memanfaatkan algoritma *Naive Bayes*. Data dikumpulkan melalui *web scraping* dalam periode 20-28 Oktober 2024 dari platform *YouTube* dan *X (Twitter)*, dengan total 2433 komentar. Proses analisis melibatkan tahapan *preprocessing*, meliputi pembersihan data, *normalisasi*, *tokenisasi*, dan *stemming* untuk meningkatkan kualitas data. Selanjutnya, fitur teks diekstraksi menggunakan metode *TF-IDF*, menghasilkan 4912 fitur. Untuk mengatasi ketidakseimbangan distribusi data sentimen, diterapkan teknik *SMOTE*, sehingga distribusi data menjadi seimbang di antara tiga Kategori sentimen yang digunakan meliputi positif, netral, dan negatif dengan jumlah kelas positif sebanyak 1332 komentar, kelas netral dengan jumlah 618 dan Sentimen negatif memiliki jumlah paling sedikit dengan jumlah 448. Berdasarkan hasil analisis, algoritma *Naive Bayes* menunjukkan bahwa algoritma *Naive Bayes* memiliki performa cukup baik dengan akurasi rata-rata 75%, *F1-score* 75%, *precision* 76%, dan *recall* 75%. Evaluasi menggunakan *confusion matrix* mengungkap bahwa model mampu mengklasifikasikan sentimen positif dengan baik, tetapi terdapat kesalahan klasifikasi pada sentimen negatif dan netral. Penelitian ini memberikan wawasan tentang pola persepsi publik terhadap isu politik tertentu, khususnya pelantikan kabinet. Diharapkan hasil ini dapat menjadi acuan untuk analisis opini masyarakat dan pengembangan model yang lebih akurat untuk analisis sentimen di masa depan

Kata Kunci: Analisis Sentimen, Media Sosial, Algoritma *Naive Bayes*.

Ini adalah artikel akses terbuka di bawah lisensi [CC BY-SA](#).



Penulis Koresponden:

Ade Krisna Subagyo

Program Studi Sistem Informasi, Fakultas Rekayasa Industri, Telkom University Purwokerto Jl. DI Panjaitan No.128, Kabupaten Banyumas Jawa Tengah, Indonesia Email: adekrisna@student.telkomuniversity.ac.id

I. PENDAHULUAN

Pelantikan kabinet baru selalu menjadi momen penting dalam sejarah politik dan pemerintahan di Indonesia. Pada tahun 2024, pelantikan Kabinet Merah Putih yang dipimpin oleh Presiden Prabowo Subianto dan Wakil Presiden Gibran Rakabuming mendapatkan berbagai reaksi dari berbagai kalangan masyarakat. Para menteri yang dilantik dalam Kabinet Merah Putih, dengan pengalaman dan tanggung jawab mereka, tentu akan menghadapi berbagai opini publik terkait kinerja mereka. Sebelum era internet, opini, kritik, dan saran Masyarakat disampaikan melalui media cetak. Namun, dengan kemajuan teknologi yang pesat, berbagai *platform* media *online*, termasuk media berita dan media sosial, kini telah menjadi sarana utama bagi masyarakat untuk menyuarakan pendapat, memberikan kritik, serta menyampaikan saran.[1]

Maraknya perkembangan media *online* yang bermunculan, masyarakat Indonesia yang dikenal dengan partisipasi aktif mereka dalam berbagai isu politik, menunjukkan antusiasme tinggi dalam merespons susunan kabinet baru. Opini publik mengenai aktivitas pemerintahan disampaikan melalui berbagai cara,

mulai dari aksi unjuk rasa hingga penggunaan platform media sosial seperti *Instagram, Facebook, Youtube, dan X* (sebelumnya *Twitter*). [2]

Komentar dan opini yang muncul di media sosial seperti *Youtube* dan *Twitter* selalu beragam karena dua platform tersebut sering dijadikan media untuk membicarakan isu yang sedang terjadi di masyarakat, [3] dan dengan banyaknya opini tersebut dapat dimanfaatkan dengan mengklasifikasikan komentar atau *tweet* untuk memperoleh informasi yang lebih terstruktur. Penelitian ini menggunakan dua sumber data, yaitu *Twitter* dan *YouTube*, untuk menganalisis sentimen dari opini tentang Kabinet Merah Putih. Dengan tujuan untuk mendapatkan data yang lebih beragam dan opini yang lebih beragam, analisis lintas platform dapat mengungkap interaksi penting yang mungkin terlewatkan jika hanya menggunakan satu platform saja. [4] Proses analisis opini atau komentar dalam bentuk teks termasuk dalam ranah analisis sentimen. Data berupa komentar atau opini diambil dari *tweet* yang kemudian diproses menggunakan metode *text mining*, yang merupakan bagian dari *data mining*. Selanjutnya, data dikategorikan sentimen, yaitu positif, negatif, dan netral. Metode ini memungkinkan analisis yang lebih mendalam untuk memahami persepsi dan respons masyarakat terhadap berbagai peristiwa atau kebijakan. [5]

Penerapan analisis sentimen dengan metode *machine learning* dapat menggunakan beberapa metode yang umum diterapkan seperti *Naive Bayes, Support Vector Machine, K-Nearest Neighbor dan Decision Tree*. [6] Pada penelitian terdahulu yang dilakukan oleh Francis Matheos Sarimole dan Wahyu Septian yang berjudul “Analisis Sentimen Masyarakat Terhadap Isu Penundaan Pemilu 2024 Pada *Twitter* Dengan Metode *Naive Bayes* Dan *Support Vector Machine*” diperoleh hasil bahwa algoritma *Naive Bayes* lebih unggul dengan akurasi 98,80% disbanding algoritma *Support Vector Machine* yang memperoleh hasil akurasi 91,61%. [7] Penelitian lainnya dilakukan oleh Muhammad Iqbal, Ade Davy Wiranata, Rayhan Suwito dan Ridha Faiz Ananda dengan judul “Perbandingan Algoritma *Naive Bayes, KNN, dan Decision Tree* terhadap Ulasan Aplikasi *Threads* dan *Twitter*” diperoleh hasil akurasi yang didapat untuk *Naive Bayes* adalah 77,79% untuk dataset *twitter* dan 65,52% untuk dataset *threads*, *Decision Tree* mendapat akurasi 72,97% untuk dataset *twitter* dan 65,43 untuk dataset *threads*, dan *KNN* mendapat akurasi 71,57 untuk dataset *twitter* dan 62,46% untuk dataset *threads*. [8] Penelitian ini menggunakan algoritma *Naive Bayes*, yang dipilih karena sifatnya yang sederhana namun efektif untuk tugas klasifikasi teks. Algoritma ini diharapkan dapat mengungkap pola-pola sentimen utama serta menganalisis perkembangan tren penyebaran informasi dan opini mengenai isu-isu tertentu di media sosial. Adanya pendekatan ini, penelitian dapat memberikan gambaran yang lebih jelas tentang respons masyarakat terhadap berbagai peristiwa, termasuk pelantikan kabinet, dan mengungkap bagaimana opini publik tersebar melalui platform media sosial. [9]

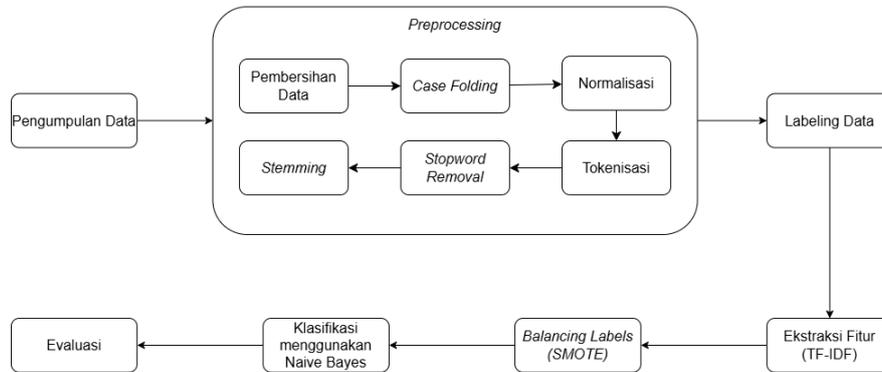
Selain metode yang digunakan, kegiatan *preprocessing* data merupakan hal yang penting saat melakukan analisis sentimen, *Data preprocessing* adalah langkah penting dalam analisis data dan *machine learning* yang bertujuan untuk mempersiapkan serta mengolah data mentah menjadi bentuk yang lebih terstruktur dan siap digunakan. Proses ini tidak hanya bertujuan meningkatkan kualitas dan akurasi data, tetapi juga memastikan data dapat dimanfaatkan secara optimal oleh model *machine learning*. Dengan melakukan *data preprocessing* secara menyeluruh, kita dapat menciptakan dataset yang berkualitas tinggi dan andal untuk analisis lebih lanjut, sekaligus mendukung kinerja model yang akan dibangun. [10]

Penelitian ini berfokus pada peran penting *preprocessing* dalam analisis sentimen menggunakan data dari *Twitter* dan *YouTube* terkait pelantikan Kabinet Merah Putih 2024. Fokus utama penelitian ini adalah mengevaluasi efektivitas tahap *preprocessing* dalam meningkatkan akurasi model klasifikasi sentimen, seperti *Naive Bayes*. Dengan demikian, tujuan akhir penelitian ini adalah untuk memberikan metode *preprocessing* yang optimal sehingga mampu meningkatkan performa model dalam mengklasifikasikan sentimen positif, negatif, atau netral secara lebih akurat. Hasil penelitian ini diharapkan dapat memberikan wawasan yang mendalam mengenai respons publik dan isu-isu yang menjadi perhatian utama masyarakat terkait pelantikan Kabinet Merah Putih 2024.

II. METODE PENELITIAN

Metode penelitian ini dilakukan melalui beberapa tahapan yang sistematis untuk memastikan kualitas dan akurasi analisis sentimen, tahap pertama dimulai dengan pengumpulan data melalui teknik *web scraping* terhadap komentar di media sosial *Youtube* dan *X* (sebelumnya *Twitter*). Teknik ini bertujuan untuk memperoleh opini masyarakat terkait kabinet merah putih. Dalam proses pengumpulan data, komentar yang diambil mencakup berbagai perspektif, sehingga memberikan Gambaran yang lebih komprehensif tentang sentimen publik. Data yang terkumpul kemudian diberi label positif, negatif, dan netral. Positif adalah teks yang mengekspresikan perasaan atau opini yang mendukung seperti “mantap buat pak Prabowo dan mas Gibran lanjutkan era Jokowi dan lebih mementingkan rakyat kecil”. Negatif adalah

teks yang menyampaikan emosi negatif atau kritik seperti “Syukur yang menjadi menteri tidak ada yang ikut anggota banteng”. Netral adalah teks yang tidak menunjukkan dukungan atau kritik terhadap suatu topik, melainkan menyampaikan informasi atau opini yang bersifat objektif, deskriptif, atau netral secara emosional, seperti “sosok Mayor Teddy pilih jadi sekretariat Kabinet”. Setelah itu dilakukan pra-pemrosesan data untuk membersihkan dan mempersiapkan teks opini hingga siap untuk diterapkan algoritma *Naive Bayes* yang digunakan untuk menganalisis sentimen serta mengelompokkan sentimen dari berbagai opini tersebut.



Gambar 1 Alur Penelitian

Tahapan Penjelasan mengenai penelitian pada Gambar 1 diatas adalah sebagai berikut:

A. Pengumpulan Data

Pengumpulan data merupakan suatu proses sistematis yang mengikuti standar prosedur untuk mendapatkan pengukuran variabel dan jawaban atas pertanyaan penelitian. Pengumpulan data penelitian ini menggunakan teknik *web scraping* untuk mendapatkan data dari media sosial *YouTube* dan *X*. Hasil *web scraping* pada media sosial *YouTube* mendapatkan 920 komentar dan *X* mendapatkan 1513 *tweet*, sehingga total data yang berhasil di dapatkan pada media sosial *YouTube* dan *X* sebanyak 2433 data. [11].

B. PreProcessing

PreProcessing merupakan tahapan persiapan data sebelum di terapkan pada model *machine learning*. Tahapan *preprocessing* ini membantu model dalam memahami data dengan lebih baik dan mengurangi kebisingan data, sehingga proses analisis menjadi lebih akurat. Terdapat beberapa tahapan pada *preprocessing* di penelitian ini seperti:

- a. *Cleaning*, Proses membersihkan data dari tanda baca, simbol, atau elemen penulisan yang tidak penting disebut pembersihan..
- b. *Case Folding*, Tahap yang disebut "*case folding*" adalah mengubah huruf kapital menjadi huruf kecil, juga disebut "*lowercase*".
- c. *Normalisasi*, *Normalisasi* merupakan tahapan mengubah kata-kata yang singkat dan tidak baku menjadi kata baku yang sesuai dengan kamus yang digunakan untuk normalisasi.
- d. *Tokenizing*, Proses membagi teks menjadi bagian-bagian kata yang terstruktur dikenal sebagai tokenisasi. Sebagai contoh, kata "saya pergi ke pasar" dibagi menjadi "saya", "pergi", "pasar", dan sebagainya.
- e. *Stopword Removal*, *Stopword removal* merupakan langkah yang akan menghilangkan kata-kata yang tidak diperlukan dalam analisis dan hanya akan memperburuk kinerja.
- f. *Stemming*, Tahap *stemming* adalah mengubah kata tambahan dari data komentar *YouTube* menjadi kata dasar. [12].

C. Labeling Data

Labeling data merupakan proses memberi kategori pada data teks berdasarkan sentimen yang terkandung di dalamnya. Dalam penelitian ini, proses labeling dilakukan menggunakan Kamus *Inset Lexicon* yang berbasis *lexicon-based*. Labeling data pada analisis sentimen ini terdiri dari tiga kategori, yaitu positif, negatif, dan netral. Metode ini membantu dalam mengotomatisasi proses labeling, sehingga model *machine learning* dapat mengidentifikasi pola-pola linguistik yang terkait dengan perasaan atau opini tertentu dengan lebih akurat [13].

D. Fitur Ekstraksi (TF-IDF)

Pembobotan TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah teknik untuk mengekstrak fitur yang bertujuan mengukur tingkat kepentingan suatu kata dalam sebuah dokumen dibandingkan dengan keseluruhan kumpulan dokumen atau *corpus* [14]. Metode ini memberikan bobot yang menunjukkan seberapa relevan suatu kata, sehingga sangat membantu dalam proses klasifikasi. Berikut adalah rumus TF-IDF :

$$TFIDF(t) = TF * \log \frac{N}{df} \quad (1)$$

Berdasarkan persamaan (1), berikut adalah keterangannya :

TF : *Term frequency*, yaitu jumlah kemunculan suatu istilah dalam sebuah komen
N : Jumlah total dokumen dalam korpus
df : Jumlah dokumen yang mengandung istilah tertentu (*document frequency*)
TFIDF(t) : Hasil perkalian antara TF dan IDF

E. Balancing Labels (SMOTE)

Balancing Labels menggunakan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) merupakan metode yang digunakan untuk mengatasi masalah kurang seimbangnya kelas atau label dalam dataset, SMOTE bekerja dengan cara *over-sampling* kelas minoritas dengan membuat contoh sintesis baru. Teknik ini membantu meningkatkan performa model *machine learning* dengan menyediakan jumlah data yang seimbang antar kelas [15].

F. Klasifikasi Menggunakan Naïve Bayes

Naïve Bayes merupakan sebuah algoritma dalam *machine learning* yang dimanfaatkan untuk mengklasifikasikan data berdasarkan teori probabilitas. Teorema Bayes digunakan dalam algoritma Naïve Bayes untuk menghitung probabilitas suatu kelas berdasarkan fitur yang ada dalam data [16]. Formula dasar teorema Bayes adalah

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (2)$$

Berdasarkan persamaan (2), berikut adalah keterangannya :

X : Data dengan class yang belum diketahui
C : Hipotesis data X merupakan suatu kelas spesifik
P(C|X) : Kemungkinan hipotesis *C* berdasarkan kondisi *X*
P(C) : Kemungkinan hipotesis *C*
P(X|C) : Kemungkinan X berdasarkan kondisi hipotesis *C*
P(X) : Kemungkinan X

G. Evaluasi

Model Evaluasi merupakan proses untuk menilai kinerja dari sebuah model pembelajaran mesin setelah dilatih pada data pelatihan. Evaluasi model penting untuk mengetahui sejauh mana model dapat melakukan prediksi dengan baik, serta untuk memilih model yang paling tepat untuk suatu masalah [17]. Beberapa metrik evaluasi yang umum digunakan dalam klasifikasi meliputi *accuracy* yang digunakan untuk mengukur sering model memberikan prediksi yang benar dibandingkan dengan semua prediksi, *precision* digunakan untuk mengukur ketepatan model dalam memprediksi kelas positif, yaitu persentase prediksi positif yang benar, *recall* digunakan untuk mengukur seberapa baik model mendeteksi semua kasus positif, yaitu persentase dari total data positif yang berhasil dikenali dengan benar, dan *f1-score* merupakan digunakan untuk mempertimbangkan kedua metrik secara bersamaan:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

Keterangan:

TP (*True Positive*) : Prediksi benar untuk kelas positif.
 TN (*True Negative*) : Prediksi benar untuk kelas negatif.
 FP (*False Positive*) : Prediksi salah untuk kelas positif.
 FN (*False Negative*) : Prediksi salah untuk kelas negatif.

III. HASIL DAN PEMBAHASAN

A. Pengumpulan data

Penelitian ini dilakukan untuk melakukan analisis sentimen masyarakat mengenai pelantikan kabinet merah putih pada media sosial *YouTube* dan *X*. Penelitian ini menggabungkan dua sumber data, yaitu *YouTube* dan *X*, untuk menganalisis sentimen dari opini tentang Kabinet Merah Putih. Dengan tujuan untuk mendapatkan data yang lebih beragam dan opini yang lebih beragam, analisis lintas *platform* dapat mengungkap interaksi penting yang mungkin terlewatkan jika hanya menggunakan satu platform saja. Data yang digunakan adalah komentar *twitter* pada tanggal 20 Oktober 2024 sampai dengan 21 Oktober 2024 dan berhasil mengumpulkan sebanyak 1513 tweet. Komentar *youtube* dari rentang waktu 21 Oktober 2024 sampai dengan 28 Oktober 2024 dan berhasil mengumpulkan data sebanyak 920 komentar.

Tabel 1 Data Hasil Web Scraping

NO	Data
1	Mhn Presiden R I.I.tentang semua produk yg mengesahkn halal dan tidak itu seharusnya MENTRI KESEHATAN.... BUKAN dari organisasi islam atau M U ICAPUT M U I tidak layak....yg layak harusnya pemerintah MENTRI KESEHATAN....BUKAN M U I... TRIMK PAK PRESIDEN PROBOWO SB.....AMIN.biar tidak disalahkan rakyat indonesia....krn M U I bukan fungsinya....m u i organisasi agama islam....tidak arti kesehatan....taunya organisasi islam indonesia...
2	Anissa AHY couple goals bagedâ€ seneng lihatnyaâ€ Terharu Mas Didiet selalu mendampingi Ayahnya ðŸ˜¸
3	Soo annoying dgn mc nya bisa ngk jn bikin bunyi mulutny sblm bicara berulang kli berisik mulutnya terlalu dekat sma mic
4	Sebagai warga negara yang baik tentunya, Pak Prabowo didampingi wakilnya dalam bekerja yang terus tetap jadi acuan Semua anggota kabinet Merah -Putih menjalankan kewajiban
2433	Prabowo-Gibran umumkan 109 orang yang akan duduk dalam susunan Kabinet Merah Putih. https://t.co/00eZGwGQZz #Foto #AdadiKompas

B. Preprocessing

Data tweet dan komentar youtube yang berhasil dikumpulkan, masuk kedalam *preprocessing*. Dalam *preprocessing* terdapat beberapa tahapan penting, meliputi pembersihan data, *case folding*, normalisasi kata, tokenisasi, *stopword removal*, dan *stemming*.

a. Pembersihan data

Pembersihan data dilakukan dengan menghapus elemen-elemen yang tidak relevan seperti angka, url, tagar, dan simbol. Hal ini bertujuan untuk meningkatkan kualitas data agar data dalam kondisi bersih, terstruktur, dan siap digunakan untuk pemodelan. Hasil pembersihan data dapat dilihat pada Tabel 2.

Tabel 2. Contoh Hasil Pembersihan Data

Sebelum Pembersihan Data	Sesudah Pembersihan Data
Mhn Presiden R I.I.tentang semua produk yg mengesahkn halal dan tidak itu seharusnya MENTRI KESEHATAN.... BUKAN dari organisasi islam atau M U ICAPUT M U I tidak layak....yg layak harusnya pemerintah MENTRI KESEHATAN....BUKAN M U I... TRIMK PAK PRESIDEN PROBOWO SB.....AMIN.biar tidak disalahkan rakyat indonesia....krn M U I bukan fungsinya....m u i organisasi agama islam....tidak arti kesehatan....taunya organisasi islam indonesia....	Mhn Presiden R Itentang semua produk yg mengesahkn halal dan tidak itu seharusnya MENTRI KESEHATAN BUKAN dari organisasi islam atau M U I CAPUT M U I tidak layak yg layak harusnya pemerintah MENTRI KESEHATANBUKAN M U I TRIMK PAK PRESIDEN PROBOWO SBAMINbiar tidak disalahkan rakyat indonesia krn M U I bukan fungsinyam u i organisasi agama islamtidak arti kesehatantaunya organisasi islam indonesia

b. Case Folding

Data yang sudah melalui pembersihan, selanjutnya dilakukan *case folding* untuk mengubah semua huruf dalam data teks menjadi huruf kecil (*lowercase*). *Case folding* menggunakan sebuah fungsi dari *python* yaitu *lower()*, fungsi ini tidak mengubah karakter lain seperti angka, spasi dan simbol. Berikut adalah contoh hasil dari *case folding* pada penelitian ini dapat dilihat pada Tabel 3.

Tabel 3 Contoh Hasil Case Folding

Sebelum Case Folding	Sesudah Case Folding
Mhn Presiden R Itentang semua produk yg mengesahkn halal dan tidak itu seharusnya MENTRI KESEHATAN	mhn presiden r itentang semua produk yg mengesahkn halal dan tidak itu seharusnya menteri kesehatan bukan

BUKAN dari organisasi islam atau M U I CAPUT M U I tidak layakg layak harusnya pemerintah MENTRI KESEHATANBUKAN M U I TRIMK PAK PRESIDEN PROBOWO SBAMINbiar tidak disalahkan rakyat indonesiaikrn M U I bukan fungsinyam u i organisasi agama islamtidak arti kesehatantaunya organisasi islam indonesia	dari organisasi islam atau m u i caput m u i tidak layakg layak harusnya pemerintah mentri kesehatanbukan m u i trimk pak presiden probowo sbaminbiar tidak disalahkan rakyat indonesiaikrn m u i bukan fungsinyam u i organisasi agama islamtidak arti kesehatantaunya organisasi islam indonesia
--	--

c. Normalisasi

Setelah melakukan *case folding* pada data, selanjutnya adalah melakukan normalisasi. Proses ini digunakan untuk mengubah kata yang tidak baku menjadi kata yang baku. Contoh hasil normalisasi pada penelitian ini dapat dilihat pada Tabel 4.

Tabel 4 Contoh Hasil Normalisasi

Sebelum Normalisasi	Sesudah Normalisasi
mhn presiden r itentang semua produk yg mengesahkn halal dan tidak itu seharusnya menteri kesehatan bukan dari organisasi islam atau m u i caput m u i tidak layakg layak harusnya pemerintah mentri kesehatanbukan m u i trimk pak presiden probowo sbaminbiar tidak disalahkan rakyat indonesiaikrn m u i bukan fungsinyam u i organisasi agama islamtidak arti kesehatantaunya organisasi islam indonesia	mohon presiden r tentang semua produk yang mengesahkan halal dan tidak itu seharusnya menteri kesehatan bukan dari organisasi islam atau m u saya caput m u saya tidak layak seharusnya pemerintah menteri m u saya bapak presiden probowo sbaminbiar tidak disalahkan rakyat indonesia m u saya bukan fungsinya u saya organisasi agama islam arti kesehatan organisasi islam indonesia

d. Tokenisasi

Data teks yang telah dilakukan normalisasi untuk mengubah kata tidak baku menjadi kata baku, selanjutnya masuk ketahap tokenisasi. Pada penelitian ini menggunakan *library* dalam python yaitu *Natural Language Toolkit* (NLTK) dan sebuah fungsi *word_tokenize* untuk memecah (tokenisasi) sebuah teks menjadi kata-kata individual. Hasil dari tokenisasi dapat dilihat pada Tabel 5.

Tabel 5 Contoh Hasil Tokenisasi

Sebelum Tokenisasi	Sesudah Tokenisasi
mohon presiden r tentang semua produk yang mengesahkan halal dan tidak itu seharusnya menteri kesehatan bukan dari organisasi islam atau m u saya caput m u saya tidak layak seharusnya pemerintah menteri m u saya bapak presiden probowo sbaminbiar tidak disalahkan rakyat indonesia m u saya bukan fungsinya u saya organisasi agama islam arti kesehatan organisasi islam indonesia	['mohon', 'presiden', 'r', 'tentang', 'semua', 'produk', 'yang', 'mengesahkan', 'halal', 'dan', 'tidak', 'itu', 'seharusnya', 'menteri', 'kesehatan', 'bukan', 'dari', 'organisasi', 'islam', 'atau', 'm', 'u', 'saya', 'caput', 'm', 'u', 'saya', 'tidak', 'layak', 'seharusnya', 'pemerintah', 'menteri', 'm', 'u', 'saya', 'bapak', 'presiden', 'probowo', 'sbaminbiar', 'tidak', 'disalahkan', 'rakyat', 'indonesia', 'm', 'u', 'saya', 'bukan', 'fungsinya', 'u', 'saya', 'organisasi', 'agama', 'islam', 'arti', 'kesehatan', 'organisasi', 'islam', 'indonesia']

e. *Stopword Removal*

Data yang telah di pecah menjadi kata-kata individual, selanjutnya dilakukan proses *stopword removal* untuk menghapus kata yang tidak penting dan akan memperburuk performa model. Pada proses ini, peneliti menggunakan modul *StopWordRemoverFactory* pada Pustaka Sastrawi. Pada Tabel 6 melihatkan data yang sudah dan sebelum dilakukan *stopword removal* menggunakan pustaka sastrawi.

Tabel 6 Contoh Hasil Stopword Removal

Sebelum <i>Stopword Removal</i>	Sesudah <i>Stopword Removal</i>
['mohon', 'presiden', 'r', 'tentang', 'semua', 'produk', 'yang', 'mengesahkan', 'halal', 'dan', 'tidak', 'itu', 'seharusnya', 'menteri', 'kesehatan', 'bukan', 'dari', 'organisasi', 'islam', 'atau', 'm', 'u', 'saya', 'caput', 'm', 'u', 'saya', 'tidak', 'layak', 'seharusnya', 'pemerintah', 'menteri', 'm', 'u', 'saya', 'bapak', 'presiden', 'probowo', 'sbaminbiar', 'tidak', 'disalahkan', 'rakyat', 'indonesia', 'm', 'u', 'saya', 'bukan', 'fungsinya', 'u', 'saya', 'organisasi', 'agama', 'islam', 'arti', 'kesehatan', 'organisasi', 'islam', 'indonesia']	mohon presiden semua produk mengesahkan halal menteri kesehatan bukan organisasi islam tidak layak seharusnya pemerintah menteri bapak presiden probowo sbaminbiar tidak disalahkan rakyat indonesia bukan fungsinya organisasi agama islam arti kesehatan organisasi islam indonesia

f. *Stemming*

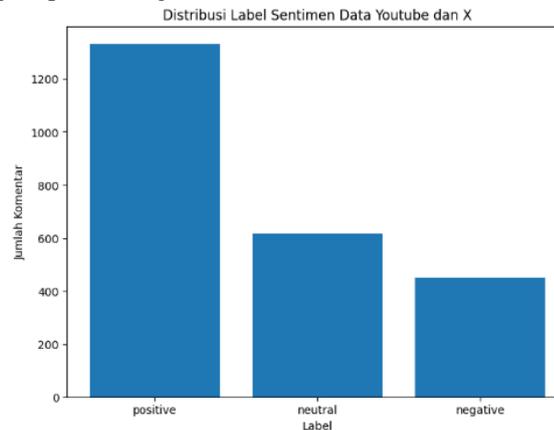
Setelah kata yang tidak penting dihilangkan, tahapan *preprocessing* yang terakhir adalah *stemming*. Pada tahap ini setiap kata akan di ubah ke bentuk dasarnya. Proses *stemming* pada penelitian ini menggunakan modul *StemmerFactory* dari pustaka *sastrawi*. Pada Tabel 6 melihat data yang sudah dan sebelum dilakukan proses *stemming* menggunakan pustaka *sastrawi*

Tabel 7 Contoh Hasil Stemming

Sebelum Stemming	Sesudah Stemming
mohon presiden semua produk mengesahkan halal menteri kesehatan bukan organisasi islam tidak layak seharusnya pemerintah menteri bapak presiden probowo sbaminbiar tidak disalahkan rakyat indonesia bukan fungsinya organisasi agama islam arti kesehatan organisasi islam indonesia	mohon presiden semua produk kesah halal menteri sehat bukan organisasi islam tidak layak harus perintah menteri bapak presiden probowo sbaminbiar tidak salah rakyat indonesia bukan fungsi organisasi agama islam arti sehat organisasi islam indonesia

C. Pelabelan Data

Pelabelan pada penelitian ini menggunakan kamus *Inset Lexicon*, kamus ini dipilih karena memiliki ribuan kata dengan polaritas yang jelas dan setiap kata dalam kamus ini memiliki bobot yang menunjukkan tingkat positif, negatif, dan netral.



Gambar 2 Distribusi Label

Grafik pada Gambar 2 diatas menunjukkan distribusi sentiment pada data yang dilakukan pelabelan menggunakan kamus *Inset Lexicon* sehingga menghasilkan tiga kategori label yaitu: positif, netral, dan negatif. Sentimen positif mendominasi dengan 1332 komentar, diikuti sentimen netral dengan jumlah 618 dan Sentimen negatif memiliki jumlah paling sedikit dengan jumlah 448. Hal ini mencerminkan respons masyarakat yang mayoritas bersifat positif terhadap topik yang dianalisis.

D. Ekstraksi Fitur

Untuk mengekstrak fitur, metode pembobotan TF-IDF (*Term Frequency-Inverse Document Frequency*) digunakan. pembobotan ini dilakukan untuk mengetahui jumlah fitur yang ada pada data penelitian. Berikut adalah hasil dari pembobotan menggunakan TF-IDF dapat dilihat pada Tabel 8.

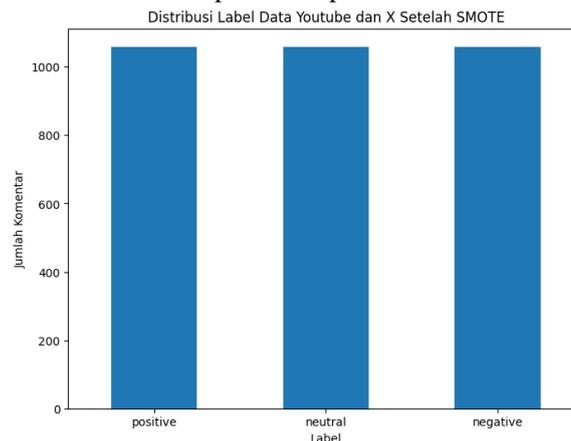
Tabel 8 Contoh Hasil TF-IDF

Indeks	aalamiin	aalhmdulilah	aalimiin	aamiin	aamiinini
0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0

Pada Tabel 8 menunjukkan hasil perhitungan *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk sejumlah istilah, yaitu *aalamiin*, *alhamdulillah*, *aalimiin*, *aamiin*, dan *aamiinini*, berdasarkan data yang dianalisis. Nilai dalam tabel merepresentasikan skor TF-IDF setiap istilah dalam dokumen tertentu.

E. Balancing Labels

Dikarenakan distribusi label pada data *YouTube* dan *X* memiliki ketidakseimbangan (*imbalanced dataset*) seperti pada Gambar 2, maka perlu dilakukan teknik *balancing* menggunakan SMOTE (*Synthetic Minority Over-sampling Technique*) untuk memastikan bahwa algoritma pembelajaran mesin (*machine learning*) tidak bias terhadap kelas mayoritas dan mampu mengenali serta memprediksi dengan baik kelas minoritas. Distribusi label data *YouTube* dan *X* setelah dilakukan teknik SMOTE dapat dilihat pada Gambar 3.



Gambar 3 Distribusi Label setelah SMOTE

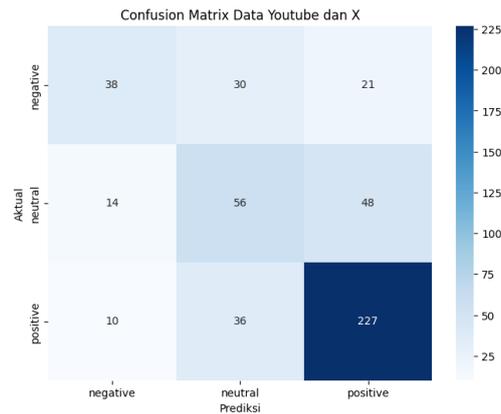
Gambar 3 menggambarkan hasil distribusi label sentimen setelah dilakukan *balancing* menggunakan SMOTE. Teknik ini berhasil menyamakan jumlah data pada setiap label (*positive*, *neutral*, *negative*) sehingga distribusi data menjadi seimbang. *Balancing* ini bertujuan untuk meningkatkan kinerja model dalam mengklasifikasikan sentimen pada kelas yang memiliki jumlah data lebih sedikit. Proses *balancing* ini dilanjutkan dengan tahap pemodelan menggunakan algoritma *Naïve Bayes*

F. Klasifikasi Naïve Bayes

Penelitian ini melakukan *modelling* menggunakan algoritma klasifikasi *Naïve Bayes*. Data yang telah dilakukan *balancing* menggunakan SMOTE diuji menggunakan *Stratified K-Fold Cross-Validation* dengan 5 lipatan (*folds*), yang menjaga proporsi distribusi kelas agar tetap seimbang di setiap lipatan. Dari hasil *cross-validation*, skor akurasi yang diperoleh pada setiap *fold* adalah 0.76, 0.76, 0.79, 0.70, dan 0.72, dengan rata-rata skor akurasi 74.98%. Standar deviasi sebesar 2.92% menunjukkan variasi performa antar lipatan relatif kecil, yang menandakan model memiliki stabilitas yang baik. Hasil ini mengidentifikasi bahwa model *Naïve Bayes* mampu mengklasifikasikan sentimen Masyarakat secara efektif berdasarkan data yang digunakan dalam penelitian ini.

G. Evaluasi Model

Evaluasi model dilakukan lebih lanjut menggunakan *confusion matrix*, yang memberikan gambaran detail performa model dalam mengklasifikasikan setiap label sentimen. Dari *confusion matrix*, dapat dianalisis jumlah prediksi benar dan salah untuk setiap kelas (*positive*, *neutral*, *negative*). Hasil menunjukkan bahwa model *Naïve Bayes* berhasil mencapai hasil yang cukup baik dengan rata-rata akurasi sebesar 75%, *F1-Score* 75%, *Precision* 76%, dan *Recall* 75%. *Confusion matrix* dapat ditampilkan pada Gambar 4.



Gambar 4 Confusion Matrix Data YouTube dan X

Berdasarkan hasil evaluasi confusion matrix, performa model dalam memprediksi sentimen berdasarkan data tersebut diperoleh hasil untuk sentimen negatif, model mampu memprediksi dengan benar 38 dari total 89 data aktual yang memiliki label negatif. Namun, terdapat beberapa kesalahan prediksi, di mana 30 data negatif diprediksi sebagai netral dan 21 data lainnya diprediksi sebagai positif. Pada sentimen netral, model berhasil memprediksi dengan benar sebanyak 56 dari total 118 data aktual yang berlabel netral. Namun, ada 14 data yang salah diklasifikasikan sebagai negatif dan 48 data salah diprediksi sebagai positif. Untuk sentimen positif, model menunjukkan performa yang paling baik, dengan memprediksi 227 dari total 273 data aktual secara akurat sebagai positif. Meskipun demikian, terdapat kesalahan pada 10 data yang diprediksi sebagai negatif dan 36 data diprediksi sebagai netral.

IV. KESIMPULAN

Penelitian Ini Menganalisis Sentimen Masyarakat Terhadap Pelantikan Kabinet Merah Putih 2024 Dengan Menggunakan Data Dari Media Sosial *Youtube* Dan *X* Yang Diklasifikasikan Menggunakan Algoritma *Naïve Bayes*. Hasil Menunjukkan Sentimen Positif Mendominasi Dengan Proporsi 55%, Diikuti Oleh Sentimen Netral Sebesar 26%, Dan Sentimen Negatif Sebesar 19%. Algoritma *Naïve Bayes* Yang Digunakan Pada Data *Youtube* Dan *X* Memiliki Performa Yang Cukup Baik Dengan Tingkat Akurasi 75%, *F1-Score* 75%, *Precision* 76%, Dan *Recall* 75%. Tahapan *Preprocessing* Seperti Pembersihan Data, Normalisasi, Dan Ekstraksi Fitur Menggunakan TF-IDF, Serta *Balancing* Dataset Dengan SMOTE, Berkontribusi Pada Peningkatan Akurasi Model Penelitian Ini Memberikan Pemahaman Mengenai Pandangan Masyarakat Terhadap Pelantikan Kabinet Merah Putih Melalui Analisis Sentimen Di Media Sosial. Hasil Ini Diharapkan Dapat Menjadi Referensi Pemerintah Dalam Memahami Opini Publik Serta Sebagai Bahan Evaluasi Untuk Meningkatkan Kepercayaan Masyarakat

REFERENSI

- [1] R. Pebrianto, S. N. Nugraha, A. Latif, and M. R. Firdaus, "Analisis Sentimen Twitter Terhadap Menteri Indonesia Dengan Algoritma Support Vector Machine Dan Naive Bayes," vol. 17, pp. 1–12, 2022.
- [2] F. Naufal Rahman, S. Lestari, and S. Tinggi Ilmu Komputer Cipta Karya Informatika Jakarta, "Analisis Sentimen Masyarakat Terhadap Pemerintah Di Era Kabinet Joko Widodo Berdasarkan Sosial Media X Menggunakan Naïve Bayes Dan K-Nearest Neighbor (Knn) Analysis Of Public Sentiment Towards The Government In The Joko Widodo Cabinet Era Based On Social Media X Using Naïve Bayes And K-Nearest Neighbor (Knn)," *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 7, no. 5, 2024.
- [3] Widia, Aqsalia Zalfa, Sari Septiana, Khoirunisa Nabila, Kurniawan Fandi "Optimasi Algoritma Naive Bayes Untuk Menganalisis Sentimen Pada Konten Pindahan Ibu Kota di Youtube" *Journal of Computer and Information Systems Ampera*, Vol. 5, No. 2, May 2024 e-ISSN: 2775-2496 .

-
- [4] A. Shevtsov, M. Oikonomidou, D. Antonakaki, P. Pratikakis, and S. Ioannidis, "What Tweets and YouTube comments have in common? Sentiment and graph analysis on data related to US elections 2020," *PLoS One*, vol. 18, no. 1 January, Jan. 2023.
- [5] F. Adi Artanto, "Analisis Sentimen Opini Publik terhadap Fenomena Bunuh Diri Mahasiswa di Twitter Menggunakan Algoritma Naive Bayes," vol. 4, no. 1, pp. 70–77, 2024.
- [6] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (Svm)," vol. 8, no. 1, pp. 147–156, 2021.
- [7] P. Guru et al., "Analisis Sentimen Masyarakat Terhadap Isu Penundaan Pemilu 2024 Pada Twitter Dengan Metode Naive Bayes Dan Support Vector Machine," *Jurnal Sains dan Teknologi*, vol. 5, no. 3, p. 2024.
- [8] 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.
- [9] A. Aliyah et al., "BRIDGE: Jurnal publikasi Sistem Informasi dan Telekomunikasi Analisis Sentimen Twitter Terhadap Tren Penyebaran Informasi Pelaku Kejahatan Menggunakan Algoritma Naives Bayes".
- [10] R. Merdiansah and A. Ali Ridha, "Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT," *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, vol. 7, no. 1, pp. 221–228, 2024.
- [11] A. R. Fadilla and P. A. Wulandari, "Literature Review Analisis Data Kualitatif: Tahap Pengumpulan Data", vol 1. Agustus 2023, Page 34-46, 2023.
- [12] D. Rifaldi, Abdul Fadlil, and Herman, "Teknik Preprocessing Pada Text Mining Menggunakan Data Tweet 'Mental Health,'" *Decode: Jurnal Pendidikan Teknologi Informasi*, vol. 3, no. 2, pp. 161–171, Apr. 2023.
- [13] A. Witanti, B. Yogyakarta Jl Raya Wates-Jogjakarta, K. Sedayu, K. Bantul, and D. Istimewa Yogyakarta, "Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (Svm)," *Jurnal Sistem Informasi dan Informatika (Simika) P-ISSN*, vol. 5, pp. 2622–6901, 2022.
- [14] F. Nur Rozi and D. Harini Sulistyawati, "Klasifikasi Berita Hoax Pilpres Menggunakan Metode Modified K-Nearest Neighbor Dan Pembobotan Menggunakan Tf-Idf," 2019.
- [15] M. Afrad, C. Febrianto, S. Wijayanto, and Y. Fathoni, "Edu Komputika Journal Sentiment Analysis of Visitor Reviews on Baturaden Tourist Attraction Using Machine Learning Methods," *Edu Komputika*, vol. 11, no. 1, 2024.
- [16] A. Faizal, A. Susilo, Y. Irawan, and D. Juardi, "Perbandingan Lexicon Based Dan Naive Bayes Classifier Pada Analisis Sentimen Pengguna Twitter Terhadap Gempa Turki Comparison Of Lexicon-Based And Naive Bayes Classifier Methods On Sentiment Analysis Of Twitter Users To The Turkey Earthquake," *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 6, no. 2, 2023.
- [17] D. Normawati and S. A. Prayogi, "Implementasi Naive Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," 2021.