

Analisis Sentimen Terhadap Film “Dirty Vote” Pada Media Sosial X dan Youtube dengan Algoritma Naive Bayes dan SVM

Kukuh Hadi Sasongko, Atiqah Meutia Hilda*

Fakultas Teknologi Industri dan Informatika, Teknik Informatika, Universitas Muhammadiyah Prof. Dr. Hamka, Jakarta, Indonesia

Email: ¹kukuh.hadi19@gmail.com, ^{2*}atiqahmeutihilda@uhamka.ac.id

Email Penulis Korespondensi: atiqahmeutihilda@uhamka.ac.id

Submitted: 29/10/2024; Accepted: 15/11/2024; Published: 15/11/2024

Abstrak– Pemilihan presiden dan waki pencalonan Gibran sebagai calon wakil presiden. Film dokumenter Dirty Vote memperdalam diskusi tersebut dengan mengungkap praktik kecurangan dan manipulasi dalam pemilu, sehingga menimbulkan kekhawatiran publik terhadap integritas pemilu. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap film "Dirty Vote" di media sosial YouTube dan Twitter (X) menggunakan algoritma Naive Bayes dan SVM. Data dikumpulkan melalui teknik crawling pada YouTube dan Twitter (X) dari tanggal 11 Februari 2024 hingga 30 Agustus 2024. Tahapan preprocessing yang dilakukan mencakup Cleansing, Transform Cases, Tokenizing, Stopwords Removal, dan Stemming. Data yang diperoleh kemudian diklasifikasikan ke dalam kategori sentimen positif dan negatif. Evaluasi model dilakukan menggunakan Confusion Matrix yang meliputi accuracy, precision, dan recall. Hasil penelitian menunjukkan variasi performa model pada kedua media sosial. Pada YouTube, algoritma Naive Bayes mencapai akurasi 81,24%, dengan precision 63,44% dan recall 100,00%, sementara SVM menunjukkan akurasi 86,94%, precision 91,62%, dan recall 65,92%. Di Twitter (X), Naive Bayes menghasilkan akurasi tertinggi sebesar 95,13%, precision 88,86%, dan recall 100,00%, sedangkan SVM mencatatkan akurasi yang sama yaitu 95,13%, dengan precision tertinggi sebesar 99,66% dan recall 87,76%. Hasil ini menunjukkan bahwa SVM lebih unggul dalam precision, sedangkan Naive Bayes memiliki recall yang konsisten tinggi. Analisis menunjukkan ketidakpuasan terhadap integritas pemilu di hampir semua platform YouTube dan Twitter (X).

Kata Kunci: Dirty Vote; Twitter(X); Youtube; Naive Bayes; SVM

Abstract–Indonesia's presidential and vice-presidential elections in 2024 sparked widespread discussion on social media, particularly regarding Gibran's candidacy as a vice presidential candidate. The documentary Dirty Vote deepened the discussion by exposing the practice of fraud and manipulation in the election, raising public concerns about the integrity of the election. This study aims to analyze public sentiment towards the Dirty Vote film on social media YouTube and Twitter (X) using Naive Bayes and SVM algorithms. Data was collected through crawling techniques on YouTube and Twitter (X) from February 11, 2024 to August 30, 2024. The preprocessing stages include Cleansing, Transform Cases, Tokenizing, Stopwords Removal, and Stemming. The data obtained is then classified into positive and negative sentiment categories. Model evaluation is done using Confusion Matrix which includes accuracy, precision, and recall. The results showed variations in model performance on both social media. On YouTube, Naive Bayes algorithm achieved 81.24% accuracy, with 63.44% precision and 100.00% recall, while SVM showed 86.94% accuracy, 91.62% precision, and 65.92% recall. On Twitter (X), Naive Bayes produced the highest accuracy of 95.13%, precision 88.86%, and recall 100.00%, while SVM recorded the same accuracy of 95.13%, with the highest precision of 99.66% and recall 87.76%. These results show that SVM is superior in precision, while Naive Bayes has a consistently high recall. The analysis showed dissatisfaction with election integrity on almost all YouTube and Twitter (X) platforms.

Keywords: Dirty Vote; Twitter(X); Youtube; Naive Bayes; SVM

1. PENDAHULUAN

Film adalah salah satu jenis konten yang sering menjadi perhatian. Selain menjadi hiburan, film menunjukkan pandangan, prinsip, dan masyarakat dimana mereka dibuat dan ditonton. "Dirty Vote" adalah salah satu film yang memicu banyak kontroversi di media sosial dan platform video yang menyoroti tentang tantangan yang dihadapi dalam menjaga integritas pada pemilihan umum. Namun, terkadang terjadi pemilu yang tidak adil dan melanggar demokrasi. Kecurangan, juga dikenal sebagai "voting kotor", adalah salah satu fenomena yang sering terjadi dan dapat mengancam integritas dan legitimasi proses demokratis [1].

Sutardara film dokumenter Dirty Vote, Dandhy Dwi Laksono, merilis film "Dirty Vote" pada 11 Februari 2024. Film dokumenter "Dirty Vote" membahas tentang sulitnya menjaga integritas pemilihan umum pada pemilu 2024. Film ini menunjukkan berbagai bentuk kecurangan yang terjadi selama proses pemilu, seperti penipuan, intimidasi pemilih, dan manipulasi hasil pemilu.[2].

Sebanyak 212,9 juta orang di Indonesia, atau sekitar 77% dari populasi menggunakan internet, sebagian besar untuk beraktivitas di media sosial. Dengan Ini ditunjukkan oleh 167 juta orang, atau sekitar 60,4% dari total pengguna internet, yang menggunakan media sosial [3]. Dengan meningkatnya jumlah pengguna media sosial, konten yang dapat diakses juga semakin banyak. Pengguna cenderung ingin menanggapi atau mendapat pengakuan dari pengguna media sosial lainnya, seperti melalui suka dan komentar. Dalam kalimat-kalimat tersebut, Ujaran kebencian adalah perbuatan tidak menyenangkan yang memiliki elemen penghinaan, pengejekkan, atau ancaman yang tidak pantas terhadap seseorang atau kelompok. Kata-kata yang merendahkan atau menyindir digunakan di setiap kalimat tanpa alasan yang kuat atau dasar yang jelas. Fenomena ini secara

tidak langsung mengubah cara orang berkomunikasi satu sama lain, dari berdiskusi di dunia nyata menjadi berdiskusi di dunia maya. Walaupun komunikasi dua arah harus terjadi, tidak jarang satu pihak hanya dapat menerima komentar, dan orang lain membuat komentar negatif atau keji tanpa mengetahui siapa yang melakukannya [4]. Menurut Pamuji (2020), ujaran kebencian yang dilakukan oleh masyarakat adalah ekspresi kekecewaan terhadap peristiwa yang terjadi di masyarakat tergantung pada latar belakangnya [5].

Dalam film tersebut, Dandhy Dwi Laksono menggambarkan betapa mudahnya penyebaran konten negatif dan berita palsu melalui media sosial, yang dapat memengaruhi opini publik secara tidak adil. Hal ini menjadi perhatian serius dalam konteks pemilihan umum, di mana informasi yang salah dapat memengaruhi hasil suara dan proses demokrasi secara keseluruhan. Oleh karena itu, penting bagi masyarakat untuk bijak dalam mengonsumsi informasi dan tidak terpengaruh oleh konten yang tidak terverifikasi di media sosial.

Media sosial sangat berpengaruh dalam perdebatan politik, mengubah cara politisi, pemerintah, dan masyarakat berkomunikasi satu sama lain. Media sosial membantu wacana politik dengan mempertahankan suara atau pandangan [6]. Media sosial tersebut juga memungkinkan adanya interaksi antara pengguna yang memiliki pandangan yang berbeda, sehingga memungkinkan terjadinya diskusi yang memperluas sudut pandang seseorang. Dengan adanya berbagai macam platform media sosial, individu dapat merasa lebih terlibat dan memiliki pengaruh dalam proses berpikir kolektif.

Dengan mengkategorikan ulasan tweet dan komentar serta banyak pengguna Aplikasi X (Twitter) dan Youtube memberikan opini yang dapat dianalisis. Aplikasi X (Twitter) adalah Twitter adalah jejaring informasi online dan juga microblogging, yang memungkinkan pengguna mengirim SMS ke seluruh dunia dengan tweet yang terdiri dari 140 karakter [7]. Ini memungkinkan pengguna untuk menyampaikan berbagai pendapat, informasi, dan kegiatan sehari-hari [8]. Dengan menggunakan YouTube, orang dapat menonton, berbagi, dan mengakses klip video secara gratis [9]. Youtube menempati urutan media sosial terpopuler yang ada di Indonesia [10]. Menurut laporan terbaru We Are Social, jumlah pengguna YouTube di Indonesia mencapai 139 juta per Oktober 2023 [11]. YouTube, salah satu platform paling populer di negara Indonesia, memungkinkan pengguna memberikan komentar, like, dan dislike terhadap video yang ditayangkan, memungkinkan penonton berinteraksi satu sama lain dan memberikan pendapat mereka tentang video tersebut. YouTube juga menjadi platform di mana pengguna dapat memberikan tanggapan terhadap konten yang mereka tonton.

Menggunakan media sosial untuk berbagi informasi dan pendapat merupakan metode yang efektif untuk mengumpulkan opini publik. Media sosial memberi penggunanya kesempatan untuk berbagi informasi dengan masyarakat sesuai keinginan mereka [12]. Selain itu, aplikasi ini memberi kesempatan bagi masyarakat untuk secara aktif berpartisipasi dalam pembentukan opini dan pandangan. Oleh karena itu, media sosial seperti Aplikasi X (Twitter) dan YouTube tidak hanya berfungsi sebagai sarana penyampaian opini, tetapi juga sebagai alat yang efektif dalam membentuk sikap dan pandangan masyarakat.

Dalam Film dokumenter "Dirty Vote, opini positif dan negatif di tweet dapat ditemukan, tetapi dengan banyak pengguna Aplikasi X (Twitter) dan Youtube, analisis manual tidak efektif. Oleh karena itu, analisis sentimen digunakan dalam penelitian ini untuk memahami perasaan masyarakat terhadap film dokumenter "Dirty Vote". Film tersebut dipilih karena relevansinya dengan konteks sosial dan politik. Film ini telah menerima banyak komentar masyarakat, termasuk dukungan dan kritik yang tajam. Oleh karena itu, sangat penting untuk melakukan analisis sentimen terhadap diskusi film yang terkait di media sosial Aplikasi X (Twitter) dan YouTube untuk memahami bagaimana pengguna merespons dan menginterpretasikan konten.

Penelitian ini penting dalam menilai sejauh mana sentimen publik terhadap film "Dirty Vote" di media sosial Twitter dan YouTube, serta faktor apa saja yang mungkin perlu diperhatikan untuk memahami dampak film tersebut. Kesimpulannya, penelitian ini membahas beberapa aspek penting terkait dengan topik penelitian, tinjauan literatur sebelumnya, serta metode yang digunakan. Metode penelitian ini mencakup pengumpulan data komentar dan ulasan tentang film "Dirty Vote" dari media sosial Twitter dan YouTube, preprocessing data, pemodelan, penilaian kinerja model, hingga penarikan hasil akhirnya.

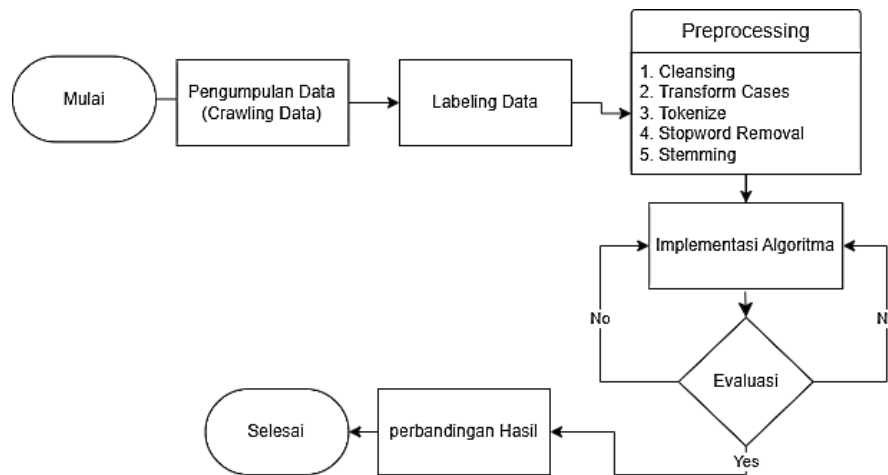
Studi sebelumnya yang dilakukan oleh Diah Fatma Sjoraida, Bucky Wibawa Karya Guna, dan Dudi Yudhokusuma (2024), menemukan bahwa tanggapan publik terhadap film "Dirty Vote" menggunakan BERT memiliki tingkat kinerja yang tinggi, dengan skor F1 di atas 0.8, dan akurasi, presisi, dan recall yang tinggi. Sentimen publik terhadap film "Dirty Vote" beragam, dengan perbedaan signifikan antara opini yang diungkapkan di media sosial (Twitter dan Facebook) dan ulasan di situs web ulasan film serta forum online. Faktor yang paling dihargai oleh masyarakat adalah efek visual dan musik, sementara pemeran dan sutradara mendapat penilaian lebih rendah [13]. Selain itu, penelitian sebelumnya yang dilakukan oleh Khairudin pada tahun 2023, berjudul Analisis Sentimen Film Di Twitter Menggunakan Metode Support Vector Machine, memperoleh akurasi sekitar 87,5% [14]. Penelitian selanjutnya, yang dilakukan oleh Rizal, berjudul Analisis Sentimen Pengguna Twitter Terkait Film One Piece Menggunakan Metode Naive Bayes, menunjukkan akurasi Naive Bayes sebesar 83,33% [15]. Studi Febriant yang lebih lanjut, berjudul Analisis Sentimen Film The Marvels Dari Aplikasi Twitter Menggunakan Metode Lexicon Based, menemukan 67,2% sentimen positif, 12,6% sentimen netral, dan 20,2% sentimen negatif terkait film The Marvels [16].

Namun pada penelitian sebelumnya tersebut hanya menggunakan metode BERT dan belum menggunakan metode lain untuk melihat dalam persepektif berbeda menggunakan metode lainnya. tujuan penelitian ini

adalah untuk melihat bagaimana masyarakat melihat film "Dirty Vote" dan menemukan hal-hal yang paling penting untuk memahami sentimen publik. Dengan menggunakan metode Lexicon Based Features untuk pelabelan data, serta analisis menggunakan algoritma Naive Bayes Classifier dan juga SVM untuk mengevaluasi hasil analisis, penelitian ini diharapkan dengan metode Naive Bayes Classifier dan SVM dapat memberikan hasil yang lebih akurat dan efisien kepada pembaca untuk mempelajari analisis film "Dirty Vote" dalam konteks sentimen sosial dan politik.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan kumpulan data dari ulasan penonton film "Dirty Vote" yang secara khusus dikumpulkan dari berbagai sumber yaitu Twitter X dan juga Youtube. **Gambar 1** menunjukkan proses penelitian secara visual, yang mencakup bagaimana peneliti merencanakan dan melakukan penelitian.



Gambar 1. Metode Penelitian

2.1 Crawling Data

Dataset yang digunakan dalam penelitian ini diperoleh melalui metode pengambilan data menggunakan metode crawling yang dijalankan pada Google Colab. Data tersebut berasal dari platform Twitter (X), data tersebut diambil menggunakan query library tweet harvest. Library Tweet Harvest mengumpulkan data dari media sosial X dengan menggunakan akses token dari akun yang sudah terdaftar di media sosial X. Sedangkan pada platform Youtube crawling data komentar dilakukan pada sebuah link ID video Youtube dan menggunakan Youtube Data API v3. Crawling Data secara efisien mengekstrak informasi yang diperlukan dari banyak dokumen [17].

2.2 Labelling Data

Setelah melakukan Crawling Data pada media sosial X dan Youtube dan mendapatkan data dari proses tersebut, setiap ulasan diberi Label sentimen yaitu label positif dan negatif. Labelling data dibagi menjadi data training dan data testing, dimana umumnya data training akan memiliki jumlah yang jauh lebih besar daripada data testing [18]. Pelabelan data dilakukan dengan menggunakan metode Lexicon based. Untuk mengetahui tanggapan atau pendapat masyarakat, Lexicon Based menggunakan kata-kata yang dievaluasi berdasarkan polaritasnya. Kata-kata ini diberi skor 1 untuk kata positif, skor -1 untuk kata negatif, dan skor 0 untuk kata netral [19].

2.3 Preprocessing Data

Preprocessing data dilakukan untuk memastikan bahwa data yang digunakan dalam model atau analisis berada dalam kondisi terbaik dan dapat memberikan hasil yang akurat [20]. Preprocessing berguna untuk membersihkan dan merapikan data. Proses ini melalui dari beberapa tahapan yaitu cleaning, transform Cases, tokenize, stop removal.

- a. Cleansing proses untuk menghapus tanda baca, angka, simbol, tautan URL, dan nama pengguna (username) dari dalam teks [21]. Data teks menjadi lebih siap untuk tahapan analisis selanjutnya, setelah tahap Cleansing.
- b. Transform Cases atau case folding adalah tahap pengolahan data teks yang bertujuan untuk mengubah semua huruf kapital (uppercase) pada dokumen diubah menjadi huruf kecil (lowercase). Tahap penting dalam proses pra-pemrosesan data linguistik. Tahap ini mencakup mengubah secara menyeluruh semua karakter alfabetis yang ada dalam dokumen atau korpus dalam format kapitalisasi yang sama [22]. Tujuan dari proses ini adalah untuk memastikan bahwa kata-kata seperti "dokumen", "Dokumen", dan "DOKUMEN" dianggap sama dalam analisis teks.

- c. Tokenisasi adalah proses pemisahan urutan karakter menjadi bagian yang lebih kecil, yang disebut token[23]. Token bisa berupa kata, frasa, atau unit teks lainnya. Proses tokenisasi dilakukan untuk menghitung bobot atau frekuensi fitur/kata pada setiap kalimat dan digunakan untuk proses klasifikasi data teks lebih lanjut. Misalnya, kalimat “demokrasi negara ini sedang tidak baik” bisa di-tokenisasi menjadi [“demokrasi”, “negara”, “ini”, “sedang”, “tidak”, “baik”].
- d. Stopword Removal adalah langkah dalam pengolahan teks yang berfungsi untuk mengurangi kata-kata yang dianggap tidak relevan atau tidak penting untuk dalam teks atau dokumen[22]. Kata-kata ini sering disebut sebagai “stopwords” dan biasanya merupakan kata-kata yang sering muncul tetapi tidak memberikan banyak informasi kontekstual, seperti “atau”, “yang”, “dengan”, “di”, “ke”, dan “tetapi”. Menghapus stopwords membantu dalam fokus pada kata-kata yang lebih penting untuk analisis, seperti analisis sentimen.
- e. *Stemming* adalah proses Mengurangi kata-kata ke bentuk dasarnya atau akarnya dengan menghilangkan awalan dan akhirnya. Tujuannya adalah untuk mengurangi jumlah kata yang perlu diproses, sehingga mempermudah analisis teks[24]. Sebagai contoh, kata "menggunakan" diubah menjadi "guna", atau "bermain" diubah menjadi "main". Proses ini membantu dalam menyederhanakan teks sehingga berbagai bentuk dari kata yang sama dianggap sebagai satu kata dasar yang sama, yang memudahkan dalam analisis dan pemrosesan teks lebih lanjut

2.4 Implementasi algoritma

Pada tahap ini, implementasi algoritma akan diterapkan pada data yang telah disiapkan. Ini adalah tahap di mana berbagai algoritma, seperti algoritma Naive Bayes dan SVM, di implementasikan dan digunakan. Naive Bayes adalah pengklasifikasian probabilitas sederhana, yang menghitung sekumpulan kemungkinan dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan [25]. SVM adalah model supervised learning yang implementasinya membutuhkan tahap pelatihan menggunakan sequential SVM dan proses pengujian setelahnya[26]. Secara khusus, metode ini digunakan untuk memperkirakan keakuratan komentar publik terhadap Film Dirty Vote yang dikategorikan sebagai positif atau negatif di YouTube dan Twitter X.

2.5 Evaluasi

Tahap ini berfungsi untuk mengevaluasi hasil dari algoritma implementasi Naive Bayes dan SVM. Sehingga diketahui Tingkat akurasi dari algoritma Naive Bayes dan SVM dalam pengklasifikasian data. Untuk mengevaluasi hasil Metode confusion matriks digunakan untuk mengevaluasi hasil prediksi, yaitu pengukuran terhadap akurasi, precision, dan recall[27].

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Gambar 2. Confusion Matrix

Pada Gambar 2. label "True Positif" dan "True Negatif" dalam kumpulan data Anda menunjukkan label pada analisis data, seperti apakah analisis data itu benar 1 atau 0, apakah data Lath benar-benar diprediksi dengan True Positif(1) atau Fasle Negatif(0), apakah data yang akan di analisis benar-benar menunjukkan prdiksi False Positif(1) atau False Negatif(0) secara akurat, dan sebagainya.

Pada label baris "Diprediksi Positif" dan "Diprediksi Negatif" menunjukkan prediksi pada model label tersebut yang di dapatkan dari data latih.

- a. *True Positives* (TP) adalah jumlah contoh positif yang dikategorikan secara akurat benar oleh model sebagai model positif.
- b. *True Negatives* (TN) adalah jumlah contoh negatif yang dikategorikan secara akurat oleh model sebagai negatif.
- c. *False Positives* (FP) adalah jumlah contoh negatif yang dikategorikan secara tidak akurat oleh model sebagai salah oleh model sebagai positif.
- d. *False Negatives* (FN) adalah jumlah contoh positif yang dikategorikan secara tidak akurat oleh model sebagai negatif.

Jumlah dokumen yang diklasifikasikan dengan benar, baik Benar Positif maupun Benar Negatif, disebut sebagai akurasi. Akurasi dapat dihitung melalui membagi jumlah data sample dengan data prediksi untuk mendapatkan jumlah akurasi yang besar. Menghitung nilai akurasi dapat menggunakan rumus[28].

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

2.6 Perbandingan Hasil

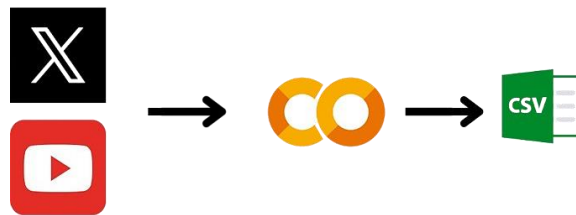
Dalam penelitian ini, analisis sentimen terhadap film “Dirty Vote” menggunakan algoritma Naive Bayes Classifier dan Support Vector Machine (SVM) menunjukkan konsistensi dalam metode pelebelan, preprocessing, pembobotan kata dengan *Lexicon Based*, serta klasifikasi sentimen. Pengumpulan data dilakukan secara terpisah antara media sosial Youtube dan Twitter X pada film tersebut dirilis, memberikan gambaran persepsi publik terhadap isu-isu yang diangkat. Metode klasifikasi Naive Bayes dan SVM digunakan secara konsisten untuk mengelompokkan tweet dan komentar YouTube menjadi sentimen positif, dan negatif. Perbandingan antara media sosial Youtube dan Twitter X dalam komentar film memungkinkan peneliti untuk mengevaluasi perbedaan sentimen masyarakat, khususnya yang terkait dengan kontroversi politik yang dipicu oleh film tersebut.

Model evaluasi menggunakan data uji dan data latih memastikan akurasi dalam mengidentifikasi sentimen di setiap media sosial, dengan metrik seperti akurasi, presisi, dan recall yang memberikan insight lebih terperinci mengenai kinerja model. Meskipun terdapat variasi pada jumlah data yang dikumpulkan pada kedua media sosial tersebut, konsistensi metode pemodelan memungkinkan analisis yang valid terhadap perbedaan sentimen publik pada Youtube dan Twitter X pada film Dirty Vote.

3. HASIL DAN PEMBAHASAN

3.1 Crawling Data

Pada tahap pengumpulan data, peneliti mendapatkan data dari Twitter (X) dan YouTube dengan menggunakan metode crawling melalui Google Colab. Kata kunci yang diambil untuk pengumpulan data melalui media sosial X yaitu “dirty vote”, pengambilan data juga mempunyai batasan yang diberikan berdasarkan bahasa Indonesia dengan query “lang:id” dan batasan waktu yaitu pada tanggal 11 Februari 2024 sampai dengan 30 Agustus 2024. Untuk pengumpulan data melalui media sosial Youtube, pengambilan data menggunakan ID link yang ada pada video serta mendaftarkan akun untuk mendapatkan layanan Youtube API v3 dan data yang diambil adalah data top level komen. Hasil pengumpulan data disimpan dalam bentuk file csv, dan jumlah komentar pada aplikasi yang dibutuhkan peneliti untuk menganalisis dicatat dalam bahasa Indonesia.



Gambar 3. Proses Crawling Data pada Twitter X dan Youtube

Pada Gambar 3 proses penarikan data ulasan menggunakan website Google Colab menggunakan kode python, Library digunakan pada google colab untuk mengakses dan menarik data di Twitter X dan youtube.

3.2 Labelling Data

Tahap berikutnya adalah pelabelan dataset yang dihasilkan dari proses crawling data. Pada tahap ini, semua data dibagi menjadi dua bagian yaitu data latih dan data uji. Data latih dilabelkan sebagai contoh untuk pengolahan data uji di masa mendatang. Untuk meningkatkan akurasi, penelitian ini hanya menggunakan data yang relevan dari data yang berhasil dikumpulkan dari internet. Dilakukan pembagian data dengan rasio 70:30, di mana 30% dari data digunakan sebagai data uji dan 70% digunakan sebagai data latih. Penelitian ini menggunakan pendekatan berbasis Lexicon untuk otomatisasi pelabelan, yang menghitung skor sentimen dari setiap teks menggunakan database kata-kata positif dan negatif. Skor ini kemudian digunakan untuk menentukan apakah komentar positif, negatif. Metode ini memungkinkan sistem untuk menganalisis dan mengklasifikasikan data secara lebih akurat dan efisien.

Tabel 1. Pelabelan Data Komentar Twitter X dengan Lexicon Based

Full_Text	Sentimen
Ketika alumni dirty vote berkumpul disitulah negara gemetar	Positive
emang bukti macam apa yg lo minta video jokowi bilang ke gibran anakku majulah jadi wapres akan	Negative

Full_Text	Sentimen
kumuluskan jalanmu aturan batas usia biar diatur paman usman ya gak mungkin ada lah tapi dari kelakuannya kan keliatan udah dijelasin gamblang di dirty vote buat yg bisa mikir	
Jawaban mudah dari raja kodok dan keluarganya emang gw pikiran Emanf bener kata closing dirty vote Ga perlu pinter cukup culas dan tahan malu Keluarga itu punya dua2nya	Negative
disodorin banyak bukti oleh ahli hukum di dokumenter dirty vote betapa sistematisnya jokowi mensukseskan anaknya menang pilpres masih bilang belum ada bukti hopeless dahhh	Positive
Ditangan orang amoral ini pilpres jd dirusak dg manipulasi suara alias curang yg ikuti DIRTY VOTE Jokowi	Negative
Akademisi bersuara itu wajar Apalagi komentar masalah kedokteran jelas masih dalam kompetensi ybs Dosen UGM ada juga yg terlibat dalam pembuatan film dirty vote yg kritik pemerintah kritiknya malah lebih parah dan ya sebagai akademisi wajar saja	Positive
Sdh dibilangin dirty vote itu kebenaran smua Eh msih di bantah terus oleh pendkung 02 Nikmatilah penderitaan ny Ok gas smpe mampus	Negative
Meloloskan anak haram konstitusi Ketua MK dipecat dari jabatannya Setelah dirty vote pilpres ketua KPU dipecat Kegilaan yang berlanjut Sama2 masalah etika dan moral	Negative
dulu siapa yg bilang dirty vote tuh cuma ingin menjatuhkan satu pihak pada realitanya memang mereka downgrade	Negative

Pada Tabel 1 menunjukkan hasil pelabelan data Twitter X. Aplikasi Twitter X dapat menghasilkan sentimen negative dan positive dengan menggunakan metode pelabelan otomatis berbasis Lexicon. Sentimen negative menunjukkan kritik terhadap kebijakan pemerintah yang berkaitan dengan film dirty vote, sedangkan sentimen positive menunjukkan penghargaan terhadap keberanian aktor yang terlibat dalam film dirty vote yang menyajikan data kecurangan.

Tabel 2. Pelabelan Data Komentar Youtube dengan Lexicon Based

full_text	sentimen
Terima kasih atas keberanian Zainal Arifin Mochtar Bivtri Susanti Feri Amsari dan seluruh team yang terlibat dalam film dokumenter Dirty Vote yang sangat edukatif amp informatif ini Terima kasih Tuhan Yesus memberkati dan tolong negri tercinta kita ini	Positive
MAU NGAKAK DIKASIH NIH VIDIO BUAT PILIH YANG CERDAS MALAH PILIH 02 SEMANGAT DIRTY VOTE SEKARANG SUDAH TERBUKTI TUNGGU SAMPAI TAHUN BERIKUTNYA DAN SETERUSNYA INDONESIA BAKALAN SUSAH	Negative
Pantes kenceng banget teriak bansos nya ngeluarin film dirty vote di masa tenang berbalut akademisi pakar hukum tatanegara ternyata anda dusta semua	Negative
TERIMAKASIH SEMUA PAKAR HUKUM dan crew SEHINGGA MEMBUAT SEBUAH FILM Dirty Vote yg sangat kami yakin bahawa kecurangan itu memang terjadi	Positive
Ditangan orang amoral ini pilpres jd dirusak dg manipulasi suara alias curang yg ikuti DIRTY VOTE Jokowi	Negative
Akademisi bersuara itu wajar Apalagi komentar masalah kedokteran jelas masih dalam kompetensi ybs Dosen UGM ada juga yg terlibat dalam pembuatan film dirty vote yg kritik pemerintah kritiknya malah lebih parah dan ya sebagai akademisi wajar saja	Positive
Ga ngaruh ternyata nih dirty vote cape cape dah bikin film	Negative
semoga semua masyarakat Indonesia tercerahkan setelah menonton dirty vote good job buat semua yang sudah terlibat dalam membuat film ini sukses selalu buat semuanya	Positive
Tanpa nonton Dirty Vote pun saya pribadi merasakan demokrasi di negara ini sedang tidak baik2 saja	Negative

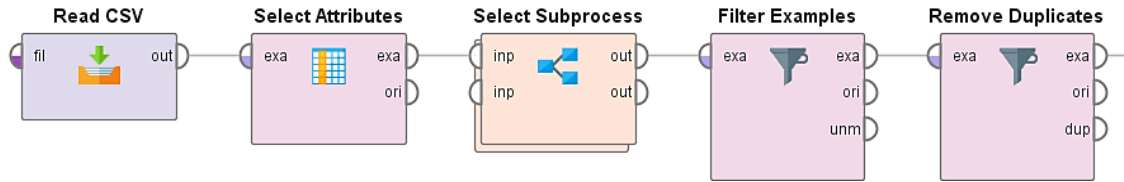
Pada Tabel 2 menunjukkan hasil pelabelan data youtube. Aplikasi youtube dapat menghasilkan sentimen negative dan positive dengan menggunakan metode pelabelan otomatis berbasis Lexicon. Sentimen negative menunjukkan kritik terhadap kebijakan pemerintah yang berkaitan dengan film dirty vote, sedangkan sentimen positive menunjukkan penghargaan terhadap keberanian aktor yang terlibat dalam film dirty vote yang menyajikan data kecurangan dan memuji kualitas video tersebut.

3.3 Preprocessing

Pada tahapan selanjutnya adalah preprocessing, Preprocessing merupakan tahapan pertama yang dilalui pada text processing, dan preprocessing ini merupakan salah satu langkah dari text mining. Preprocessing berguna untuk membersihkan dan merapikan data. Setelah hasil data crawling dari komentar aplikasi X dan YouTube selesai, tahap preprocessing dilakukan untuk teks yang digunakan sebagai input dari algoritma yang digunakan dalam penelitian ini. Setiap tahap persiapan terdiri dari sejumlah langkah yaitu, cleansing, Transform Cases, Tokenize, Stopword Removal, Stemming.

3.1.1 Cleansing

Proses Cleansing dilakukan untuk menghilangkan tanda baca, angka, simbol, tautan URL, dan nama pengguna atau hastag dari teks. Tujuannya adalah untuk menghasilkan komentar yang bersih dan relevan, yang akan memudahkan analisis, dan yang akan memastikan bahwa informasi yang diberikan sudah rapih.



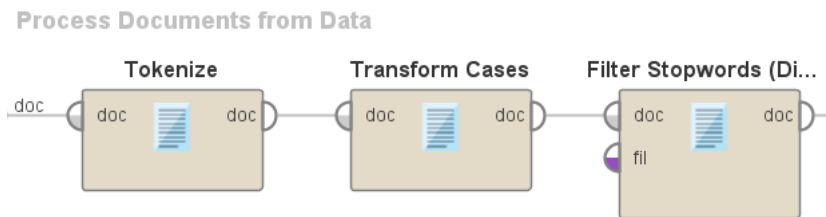
Gambar 4. Tahapan Cleansing Data

Pada Gambar 4 menunjukkan tahapan cleansing yang bertujuan untuk menghilangkan kata-kata yang sama dan simbol-simbol yang tidak diperlukan selama proses analisis pada mesin. Karena pada data komentar terdapat simbol seperti [-!"#\$%&'()*+,-./:;<=>?@\[\]_`{|}~], URL, dan juga emot dalam komentar.

Tabel 3. Hasil Cleansing

Sebelum Cleansing	Setelah Cleansing
Ya Allah mas uceng yg di film dirty vote ngisi mentoring alter IRI TOTAL GUA KEBAGIANNYA GELOMBANG 1 ARGHHH https://t.co/9deJFrXF7j @Dennysiregar7 Kan dah di jelaskan di film dokumenter dirty vote mental nya cuma culas dan tahan malu	Ya Allah mas uceng yg di film dirty vote ngisi mentoring alter IRI TOTAL GUA KEBAGIANNYA GELOMBANG 1 ARGHHH Kan dah di jelaskan di film dokumenter dirty vote mental nya cuma culas dan tahan malu

Pada Tabel 3 menampilkan hasil *cleansing* data yang menghilangkan simbol-simbol yang tidak diperlukan serta url dan emot dari komentar.



Gambar 5. Tahapan Preprocessing Data

Pada Gambar 5 Setelah data di *cleansing* data akan di proses menjadi *Transform Case*, *tokenize*, dan juga Filter Stopword untuk menyaring kata-kata hingga dapat diproses untuk analisis.

3.1.2 Transform Cases

Bagian penting dari proses pra-pemrosesan data linguistik adalah transformasi case, juga dikenal sebagai case folding, di mana semua huruf kapital (uppercase) dalam dokumen diubah menjadi huruf kecil (lowercase). Mengubah secara menyeluruh semua karakter alfabetis yang ada dalam dokumen.

Tabel 4. Transform Cases

Sebelum Transform	Setelah Transform
Ya Allah mas uceng yg di film dirty vote ngisi mentoring alter IRI TOTAL GUA KEBAGIANNYA GELOMBANG 1 ARGHHH	'ya', 'allah', 'mas', 'uceng', 'yg', 'di', 'film', 'dirty', 'vote', 'ngisi', 'mentoring', 'alter', 'iri', 'total', 'gua', 'kebagiannya', 'gelombang', '1', 'arghhh'
Kan dah di jelaskan di film dokumenter dirty vote mental nya cuma culas dan tahan malu	'kan', 'dah', 'di', 'jelaskan', 'di', 'film', 'dokumenter', 'dirty', 'vote', 'mental', 'nya', 'cuma', 'culas', 'dan', 'tahan', 'malu'

Pada Tabel 4. Menampilkan hasil *Transform Case* data yang mengubah huruf kapital dalam dokumen menjadi huruf kecil.

3.1.3 Tokenize

Konsep tokenisasi mencakup prosedur sederhana di mana urutan teks dibagi menjadi bagian yang lebih kecil. Token bisa berupa kata, frasa, atau unit teks lainnya. Proses tokenisasi proses yang digunakan untuk menghitung bobot atau frekuensi fitur atau kata yang ada di setiap kalimat. Proses ini juga digunakan dalam proses klasifikasi data teks lainnya.

Tabel 5. Tokenize

Sebelum Tokenize	Setelah Tokenize
Ya Allah mas uceng yg di film dirty vote ngisi mentoring alter IRI TOTAL GUA KEBAGIANNYA GELOMBANG 1 ARGHHH	'ya', 'allah', 'mas', 'uceng', 'yg', 'di', 'film', 'dirty', 'vote', 'ngisi', 'mentoring', 'alter', 'iri', 'total', 'gua', 'kebagiannya', 'gelombang', '1', 'arghhh'
Kan dah di jelaskan di film dokumenter dirty vote mental nya cuma culas dan tahan malu	'kan', 'dah', 'di', 'jelaskan', 'di', 'film', 'dokumenter', 'dirty', 'vote', 'mental', 'nya', 'cuma', 'culas', 'dan', 'tahan', 'malu'

Pada Tabel 5. Menampilkan hasil data Tokenize yang membagi kata menjadi bagian yang terpisah lebih kecil dalam dokumen.

3.1.4 Stopword Removal

Salah satu langkah dalam pengolahan teks adalah stopwords removal, yang menghilangkan kata-kata yang dianggap tidak penting atau tidak relevan untuk analisis lebih lanjut. Fokus pada kata-kata yang lebih penting untuk analisis, seperti analisis sentimen, dapat ditingkatkan dengan mengurangi kata-kata yang sering muncul tetapi tidak memberikan informasi yang cukup kontekstual, seperti "atau", "yang", "dengan", "di", "ke", dan "tetapi".

Tabel 6. Stopword Removal

Sebelum Stopword	Setelah Stopword
Ya Allah mas uceng yg di film dirty vote ngisi mentoring alter IRI TOTAL GUA KEBAGIANNYA GELOMBANG 1 ARGHHH	'allah', 'mas', 'uceng', 'film', 'dirty', 'vote', 'ngisi', 'mentoring', 'alter', 'iri', 'total', 'gua', 'kebagiannya', 'gelombang', '1', 'arghhh'
Kan dah di jelaskan di film dokumenter dirty vote mental nya cuma culas dan tahan malu	'jelaskan', 'film', 'dokumenter', 'dirty', 'vote', 'mental', 'culas', 'tahan', 'malu'

Pada Tabel 6. Menampilkan hasil data *Stopword Removal* yang menghilangkan kata sambung dalam dokumen.

3.1.5 Stemming

pengurangan kata-kata ke bentuk dasarnya atau akarnya dengan menghilangkan awalan dan akhirnya. Tujuan dari proses ini adalah untuk mengurangi jumlah kata harus diproses, yang berarti analisis teks menjadi lebih mudah

Tabel 7. Stemming

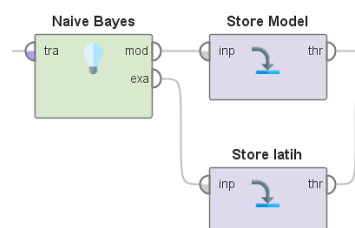
Sebelum Stemming	Setelah Stemming
Ya Allah mas uceng yg di film dirty vote ngisi mentoring alter IRI TOTAL GUA KEBAGIANNYA GELOMBANG 1 ARGHHH	'allah', 'mas', 'uceng', 'film', 'dirty', 'vote', 'ngisi', 'mentoring', 'alter', 'iri', 'total', 'gua', 'bagi', 'gelombang', '1', 'arghhh'
Kan dah di jelaskan di film dokumenter dirty vote mental nya cuma culas dan tahan malu	'jelas', 'film', 'dokumenter', 'dirty', 'vote', 'mental', 'culas', 'tahan', 'malu'

Pada Tabel 7. Menampilkan hasil data Stemming yang menghilangkan awalan dan akhiran kata dalam dokumen.

3.4 Implementasi Algoritma

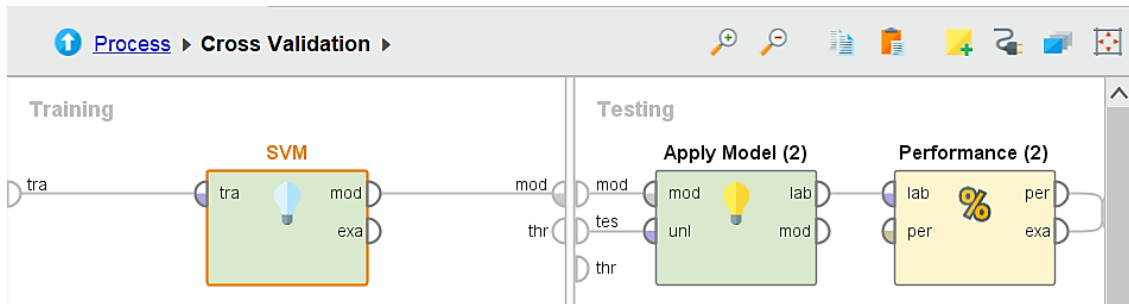
Setelah preprocessing data, implementasi algoritma sentimen adalah langkah berikutnya. Implementasi algoritma Naive Bayes dan juga SVM penambangan data tujuan dari tahap ini. Pelatihan dan implementasi berbagai algoritma penambangan data dilakukan pada tahap ini.

- Pada tahap pemodelan Naive Bayes, operator Naive Bayes disambungkan ke dua *Store* yang pertama digunakan untuk data model dan yang kedua untuk data latih, seperti yang ditunjukkan pada Gambar 6.



Gambar 6. Operator Tahap Model Naive Bayes

- b. Pada tahap pemodelan SVM berbeda dengan Naïve Bayes menggunakan operator *Cross Validation*, operator SVM disambungkan ke model dan selanjutnya model disambungkan ke Performance seperti yang ditunjukkan pada Gambar 7.

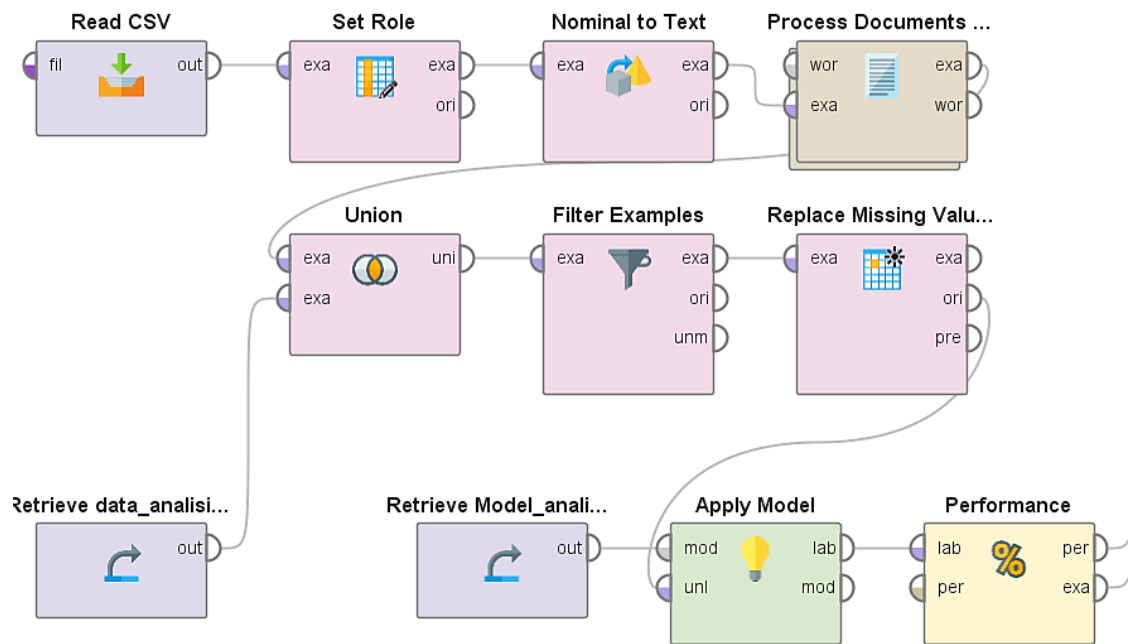


Gambar 7. Operator Tahap Model SVM

3.5 Evaluasi

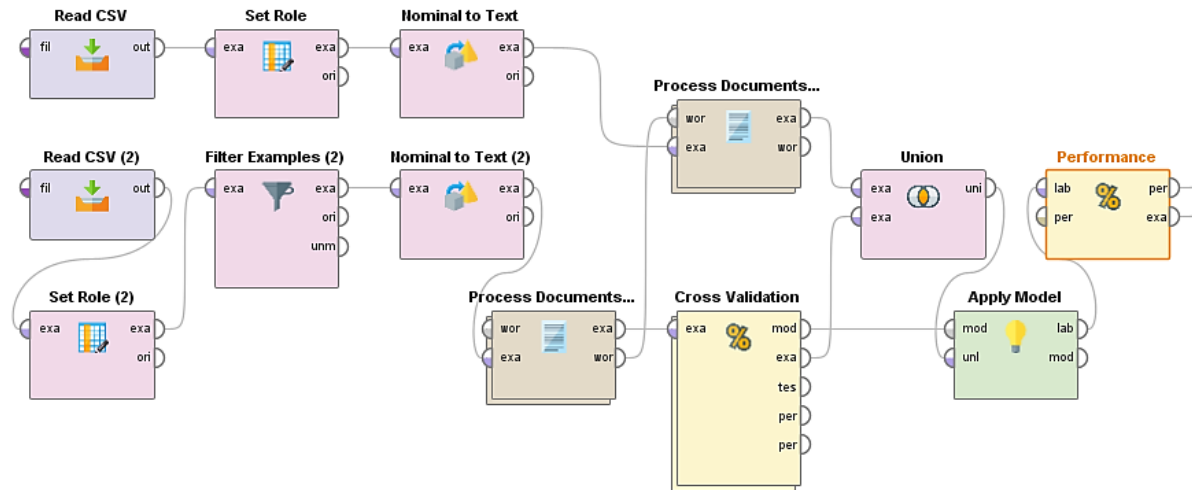
Setelah pengujian selesai, tahap berikutnya adalah evaluasi hasil, yang mengevaluasi hasil dengan menggunakan tiga parameter: *precision*, *accuray*, dan *recall*.

- a. Data dianalisis dengan metode Naïve Bayes, data pada model dan latih di hitung dalam *confusion matrix* pada aplikasi *RapidMiner* untuk mengetahui *precision*, *accuray*, dan *recall* Proses *Confusion Matrix* menggunakan aplikasi *Rapidminer*. Konten “*Read CSV*” digunakan untuk membaca data dalam file Excel dengan format CSV data tersebut adalah data uji. Operator “*Process Document*” digunakan untuk *preprocessing*. Adapun operator union tersebut adalah menggabungkan data latih dengan data uji. Kemudian model data di gabungan untuk mengetahui hasil *confusion matrix*, operator performance tersebut yang menghitung berapa jumlah hasil dari *precision*, *accuray*, dan *recall* yang di tunjukkan Pada Gambar 8.



Gambar 8. Operator perhitungan *Confusion Matrix* Naïve Bayes

- b. Data dianalisis dengan metode SVM, data model dan latihan hitung dalam matrix confusion pada aplikasi *RapidMiner* untuk mengevaluasi *precision*, *accuray*, dan *recall* Proses *Confusion Matrix* menggunakan aplikasi *Rapidminer*. Konten Konten “*Read CSV*” digunakan untuk membaca data dalam file Excel dengan format CSV data tersebut adalah data uji. Operator “*Process Document*” digunakan untuk *preprocessing*. Adapun operator union tersebut adalah menggabungkan data latih dengan data uji. Kemudian model data di gabungan untuk mengetahui hasil *confusion matrix*, penelitian dilakukan sebanyak 10 kali dengan bantuan Operator *cross validation*, operator performance tersebut yang menghitung berapa jumlah hasil dari *precision*, *accuray*, dan *recall* yang di tunjukkan Pada Gambar 9.



Gambar 9. Operator perhitungan Confusion Matrix SVM

3.6 Perbandingan Hasil

Setelah *evaluasi* selesai, tahap berikutnya adalah Perbandingan hasil, yang membandingkan hasil dengan menggunakan parameter *precision*, *accuracy*, *recall* dan juga menggunakan *True positive* dan *True Negatif*, *False positive* dan *False Negative*. dilakukan pengujian dengan Confusion Matrix yang diperoleh dari proses pemodelan Naïve Bayes dan SVM. metode dilakukan dengan pembagian data 70:30. Oleh karena itu, metode naïve bayes dengan dataset yang digunakan berjumlah 9247 pada aplikasi *Youtube*, terdapat 8430 data bersih, 6472 data latih, dan 1958 data uji, sedangkan pada aplikasi *twitter* terdapat 1788 dataset terdiri dari 1232 data bersih, 862 data latih dan 369 data uji.

Tabel 8. Confusion Matrix dari platform X, YT dan juga algoritma Naïve Bayes dan SVM

	TN	FN	FP	TP
X_Naïve Bayes	485	0	42	335
X_SVM	526	41	1	294
YT_Naïve Bayes	3151	0	1214	2107
YT_SVM	4238	718	127	1389

Di Tabel 8, *precision*, *accuracy*, *recall*, rata-rata ditampilkan dengan perhitungan menggunakan rumus pada (1), (2), (3).

Tabel 9. Confusion Matrix dengan recall, accuracy, Precision di X dan Naïve Bayes

	True. Negative	True. Positive	Class Precision
Pred. Negative	485	0	100.00%
Pred. Positive	42	335	88.86%
Class Recall	92.03%	100.00%	

Perhitungan *accuracy* dari Tabel 9. Dengan rumus *accuracy* sebagai berikut.

$$Accuracy = \frac{335 + 485}{42 + 335 + 485 + 0} = \frac{820}{862} = 95.13\%$$

Berdasarkan perhitungan dari Tabel 9 menampilkan hasil perhitungan menggunakan algoritma *Naïve Bayes*, yang menghasilkan *accuracy* 95.13% dari aplikasi X.

Tabel 10. Confusion Matrix dengan recall, accuracy, Precision di X dan SVM

	True. Negative	True. Positive	Class Precision
Pred. Negative	526	41	92.77%
Pred. Positive	1	294	99.66%
Class Recall	99.81%	87.76%	

Perhitungan *accuracy* dari Tabel 10. Dengan rumus *accuracy* sebagai berikut.

$$Accuracy = \frac{294 + 526}{294 + 41 + 526 + 1} = \frac{820}{862} = 95.13\%$$

Berdasarkan perhitungan dari Tabel 10 menampilkan hasil perhitungan menggunakan algoritma SVM, yang menghasilkan *accuracy* 95.13% dari aplikasi X.

Tabel 11. *Confusion Matrix* dengan *recall*, *accuracy*, *Precision* di Youtube dan Naïve Bayes

	<i>True. Negative</i>	<i>True. Positive</i>	<i>Class Precision</i>
<i>Pred. Negative</i>	3151	0	100.00%
<i>Pred. Positive</i>	1214	2107	63.44%
Class Recall	72.19%	100.00%	

Perhitungan *accuracy* dari Tabel 11. Dengan rumus *accuracy* sebagai berikut.

$$Accuracy = \frac{2107 + 3151}{2107 + 1214 + 3151 + 0} = \frac{5258}{6472} = 81.24\%$$

Berdasarkan perhitungan dari Tabel 11 menampilkan hasil perhitungan menggunakan algoritma Naïve Bayes, yang menghasilkan *accuracy* 81.24% dari aplikasi Youtube.

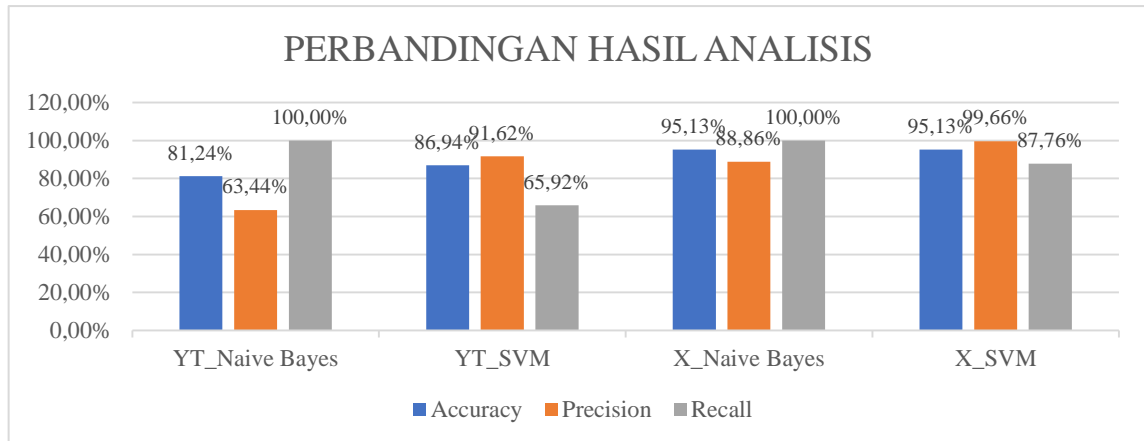
Tabel 12. *Confusion Matrix* dengan *recall*, *accuracy*, *Precision* di Youtube dan SVM

	<i>True. Negative</i>	<i>True. Positive</i>	<i>Class Precision</i>
<i>Pred. Negative</i>	4238	718	85.51%
<i>Pred. Positive</i>	127	1389	91.62%
Class Recall	97.09 %	65.92%	

Perhitungan *accuracy* dari Tabel 12. Dengan rumus *accuracy* sebagai berikut.

$$Accuracy = \frac{1389 + 4238}{1389 + 127 + 4238 + 718} = \frac{5627}{6472} = 86.94\%$$

Berdasarkan perhitungan dari Tabel 12 menampilkan hasil perhitungan menggunakan algoritma SVM, yang menghasilkan *accuracy* 86.94% dari aplikasi Youtube.



Gambar 10. Perbandingan Hasil Analisis

Pada Gambar 10. menunjukkan perbandingan hasil analisis berdasarkan confusion matrix menunjukkan bahwa dataset dari kedua media sosial menggunakan algoritma Naive Bayes dan SVM menunjukkan bahwa kinerja model berbeda-beda tergantung pada karakteristik dataset yang digunakan. Ini menunjukkan bahwa kedua algoritma memiliki kemampuan untuk membuat prediksi yang akurat

Berdasarkan hasil analisis sentimen terhadap film *Dirty Vote* di media sosial YouTube dan Twitter (X), algoritma Naïve Bayes dan SVM menunjukkan kinerja yang bervariasi. Pada media sosial YouTube, Naïve Bayes memiliki akurasi sebesar 81,24%, precision 63,44%, dan recall 100,00%, sedangkan SVM menunjukkan akurasi 86,94%, precision 91,62%, dan recall 65,92%. Sementara itu, pada Twitter (X), Naïve Bayes menghasilkan akurasi 95,13%, precision 88,86%, dan recall 100,00%, sementara SVM menunjukkan akurasi yang sama yaitu 95,13%, precision 99,66%, dan recall 87,76%. Dari hasil ini, dapat disimpulkan bahwa pada kedua platform, algoritma SVM memiliki kinerja precision yang lebih tinggi dibandingkan Naïve Bayes, terutama pada Twitter (X) dengan precision tertinggi sebesar 99,66%. Meskipun Naïve Bayes unggul dalam recall pada kedua platform, akurasi terbaik diperoleh dari algoritma Naïve Bayes dan SVM pada Twitter (X) dengan nilai akurasi yang sama yaitu 95,13%. Hal ini menunjukkan bahwa algoritma SVM dan Naïve Bayes keduanya mampu memberikan prediksi yang akurat, namun SVM terbukti lebih efektif dalam menghasilkan precision yang lebih tinggi, khususnya di platform sosial Twitter(X).

4. KESIMPULAN

Berdasarkan hasil penelitian menggunakan algoritma Naïve Bayes dan SVM pada data dari di platform sosial Twitter YouTube dan Twitter (X), diperoleh variasi dalam performa model. Pada media sosial YouTube, Naïve Bayes menunjukkan akurasi sebesar 81,24%, dengan precision 63,44% dan recall 100,00%, sedangkan SVM mencapai akurasi 86,94%, precision 91,62%, dan recall 65,92%. Sementara itu, pada media sosial Twitter (X), Naïve Bayes mencapai akurasi 95,13%, precision 88,86%, dan recall 100,00%, sementara SVM menunjukkan hasil akurasi yang sama sebesar 95,13%, dengan precision 99,66% dan recall 87,76%. Hasil penelitian menunjukkan bahwa akurasi model bervariasi antara 81,24% hingga 95,13%, dengan akurasi tertinggi diperoleh pada algoritma Naïve Bayes dan SVM pada di platform sosial Twitter (X). Nilai precision tertinggi diperoleh pada SVM di Twitter (X) dengan 99,66%, sedangkan precision terendah ada pada Naïve Bayes di YouTube dengan nilai 63,44%. Meskipun precision bervariasi, recall tetap tinggi untuk Naïve Bayes pada kedua platform dengan nilai 100,00%, menggambarkan kemampuan model untuk mendeteksi sentimen dengan akurat. Analisis menunjukkan ketidakpuasan terhadap integritas pemilu di hampir semua platform YouTube dan Twitter (X), mencerminkan reaksi publik yang negatif terhadap isu yang diangkat dalam film *Dirty Vote*. Hasil penelitian ini juga menunjukkan bahwa algoritma SVM memiliki keunggulan dalam precision, terutama pada media sosial Twitter (X). Namun, Naïve Bayes menunjukkan performa recall yang konsisten pada kedua media sosial. Temuan ini menunjukkan bahwa SVM lebih efektif dalam menghasilkan precision yang lebih tinggi, sedangkan Naïve Bayes mampu mempertahankan recall yang maksimal. Untuk meningkatkan performa model lebih lanjut, disarankan agar pengembangan model mempertimbangkan optimasi pada precision untuk memastikan identifikasi sentimen yang lebih akurat.

REFERENCES

- [1] A. A. Firdaus, A. Yudhana, I. Riadi, and Mahsun, "Indonesian presidential election sentiment: Dataset of response public before 2024," *Data Brief*, vol. 52, 2024, doi: 10.1016/j.dib.2023.109993.
- [2] A. Nathaniella and I. Triadi, "Pengaruh Film Dokumenter 'Dirty Vote' pada Saat Masa Tenang Pemilihan Umum Tahun 2024 di Indonesia," *Indonesian Journal of Law and Justice*, vol. 1, no. 4, p. 11, 2024, doi: 10.47134/ijlj.v1i4.2402.
- [3] S. Kemp, "The State of Digital in Indonesia in 2023," datareportal. [Online]. Available: <https://datareportal.com/reports/digital-2023-indonesia>
- [4] F. L. Sa'idah, D. E. Santi, and S. Suryanto, "Faktor Produksi Ujaran Kebencian melalui Media Sosial," *Jurnal Psikologi Perseptual*, vol. 6, no. 1, 2021, doi: 10.24176/perseptual.v6i1.5144.
- [5] E. Pamuji, "Ujaran kebencian pada ruang – ruang digital," *Jurnal Kajian Media*, vol. 4, no. 2, 2020, doi: 10.25139/jkm.v4i2.2811.
- [6] H. J. Sitorus, M. Tanoyo, and . I., "Polarisasi Politik Melalui Interaksi Sosial Di Instagram: Studi Kasus Pemilu 2024 Di Indonesia," *JKOMDIS: Jurnal Ilmu Komunikasi Dan Media Sosial*, vol. 4, no. 2, pp. 383–394, 2024, doi: 10.47233/jkomdis.v4i2.1675.
- [7] P. Fitriyah and A. Muhammad F, "PENENTUAN MODULARITY CLASS PADA FENOMENA CROSS PLATFORM #WHATSAPDOWN TRENDING DI TWITTER MENGGUNAKAN SOCIAL NETWORK ANALYSIS," *BroadComm*, vol. 3, no. 1, 2021, doi: 10.53856/bcomm.v3i1.216.
- [8] I. Utami and M. Marzuki, "Analisis Sistem Informasi Banjir Berbasis Media Twitter," *Jurnal Fisika Unand*, vol. 9, no. 1, 2020, doi: 10.25077/jfu.9.1.67-72.2020.
- [9] L. Sari, "Upaya Menaikkan Kualitas Pendidikan Dengan Pemanfaatan Youtube Sebagai Media Ajar Pada Masa Pandemi Covid-19," *Jurnal Tawadhu*, vol. 4, no. 1, 2020.
- [10] R. Kurniawan, F. Lestari, A. S. Batubara, M. Z. A. Nazri, K. Rajab, and R. Munir, "Indonesian Lexicon-Based Sentiment Analysis of Online Religious Lectures Review," in *2021 International Congress of Advanced Technology and Engineering, ICOTEN 2021*, 2021. doi: 10.1109/ICOTEN52080.2021.9493530.
- [11] D. A. Riyanto, "We Are Social Indonesia Digital Report 2023," *We Are Social-Hootsuite*, 2023.
- [12] F. Jamilah and P. Wahyuni, "Ujaran Kebencian dalam Kolom Komentar YouTube pada Tahun Politik Pemilihan Presiden 2019," *Silampari Bisa: Jurnal Penelitian Pendidikan Bahasa Indonesia, Daerah, dan Asing*, vol. 3, no. 2, pp. 325–341, 2020, doi: 10.31540/silamparibisa.v3i2.1109.
- [13] D. F. Sjoraida, B. W. K. Guna, and D. Yudhakusuma, "Analisis Sentimen Film Dirty Vote Menggunakan BERT (Bidirectional Encoder Representations from Transformers)," *Jurnal JTJK (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 8, no. 2, pp. 393–404, 2024, doi: 10.35870/jtik.v8i2.1580.
- [14] M. Khairudin, A. Sukendar, and A. Somantri, "ANALISIS SENTIMEN FILM DI TWITTER MENGGUNAKAN METODE SUPPORT VECTOR MACHINE," *Jurnal Sains dan Sistem Teknologi Informasi*, vol. 5, no. 1, 2023, doi: 10.59811/sandi.v5i1.47.
- [15] Muhammad Rizal, M. Martanto, and U. Hayati, "ANALISIS SENTIMEN PENGGUNA TWITTER TERKAIT FILM ONE PIECE MENGGUNAKAN METODE NAIVE BAYES," *Jurnal Sistem Informasi Kaputama (JSIK)*, vol. 8, no. 1, 2024, doi: 10.59697/jsik.v8i1.522.

- [16] F. Febriant, H. Christy, and A. Wijaya, "Analisis Sentimen Film The Marvels Dari Aplikasi Twitter Menggunakan Metode Lexicon Based," *JuSiTik : Jurnal Sistem dan Teknologi Informasi Komunikasi*, vol. 7, no. 1, 2023, doi: 10.32524/jusitik.v7i1.1046.
- [17] S. Samsir, A. Ambiyar, U. Verawardina, F. Edi, and R. Watrianthos, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 1, 2021, doi: 10.30865/mib.v5i1.2580.
- [18] W. Armadianti *et al.*, "Analisis Sentimen Netizen Terhadap Personal Branding Elon Musk Pada Platform X Dengan Pendekatan Analisis Support Vector Machine," vol. 9, no. 1, 2024.
- [19] D. Hernikawati, "Kecenderungan Tanggapan Masyarakat Terhadap Vaksin Sinovac Berdasarkan Lexicon Based Sentiment Analysis," *IPTEK-KOM*, vol. 23, no. 1, 2021.
- [20] R. Firdaus *et al.*, "Implementasi Algoritma Random Forest Untuk Klasifikasi Pencemaran Udara di Wilayah Jakarta Berdasarkan Jakarta Open Data," vol. 14, no. 2, pp. 520–525, 2021.
- [21] A. Gaizka, A. R. Dzikrillah, and E. Sinduningrum, "Analisis Sentimen Masyarakat Sebelum Dan Sesudah Terpilihnya Gibran Sebagai Cawapres Prabowo Menggunakan Naïve Bayes," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 6, pp. 2830–2841, 2024, doi: 10.30865/klik.v4i6.1876.
- [22] D. Prtamtanto *et al.*, "Analisa Sentimen Persepsi Masyarakat Terhadap Aplikasi Bea Cukai Mobile Menggunakan Algoritma Naive Bayes Dan K-Nearest," vol. 12, no. 2, pp. 92–100, 2024.
- [23] R. Nurlaely, S. D. Sartika, Kamdan, and I. L. Kharisma, "Analisis Sentimen Twitter Terhadap Cyberbullying Menggunakan Metode Support Vector Machine (SVM)," *Jurnal Computer Science and Information Technology (CoSciTech)*, vol. 4, no. 2, pp. 376–384, 2023, [Online]. Available: <http://ejournal.umri.ac.id/index.php/coscitech/indexhttps://doi.org/10.37859/coscitech.v4i2.5161>
- [24] R. Ulgasesa, A. B. P. Negara, and T. Tursina, "Pengaruh Stemming Terhadap Performa Klasifikasi Sentimen Masyarakat Tentang Kebijakan New Normal," *Jurnal Sistem dan Teknologi Informasi (JustIN)*, vol. 10, no. 3, 2022, doi: 10.26418/justin.v10i3.53880.
- [25] D. Rezekika, "Penerapan Algoritma Naïve Bayes Untuk Memprediksi Penjualan Spare Part Sepeda Motor," *Jurnal Pelita Informatika*, vol. 8, no. 3, 2020.
- [26] M. Muhathir, M. H. Santoso, and D. A. Larasati, "Wayang Image Classification Using SVM Method and GLCM Feature Extraction," *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, vol. 4, no. 2, 2021, doi: 10.31289/jite.v4i2.4524.
- [27] 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, 2021, doi: 10.29207/resti.v5i4.3146.
- [28] M. Syarifuddin, "ANALISIS SENTIMEN OPINI PUBLIK TERHADAP EFEK PSBB PADA TWITTER DENGAN ALGORITMA DECISION TREE, KNN, DAN NAÏVE BAYES," *INTI Nusa Mandiri*, vol. 15, no. 1, 2020, doi: 10.33480/inti.v15i1.1433.