

ANALISIS SENTIMEN KOMENTAR YOUTUBE TERHADAP WAWANCARA PRESIDEN PRABOWO MENGGUNAKAN MACHINE LEARNING DAN ORANGE DATA MINING

Nama : Kristina E. V. Sihombing

NIM : 672022337

02 May, 2024

Abstrak

Penelitian ini menganalisis sentimen komentar publik pada video wawancara Presiden Prabowo Subianto di YouTube. Data dikumpulkan melalui YouTube API dan diproses dengan text preprocessing serta pelabelan semi-supervised menggunakan Orange Data Mining. Tiga algoritma machine learning diterapkan: Naïve Bayes, Random Forest, dan SVM.

Hasil menunjukkan bahwa SVM memberikan performa terbaik dengan akurasi 74%, disusul Random Forest 73% dan Naïve Bayes 66%. Temuan ini menunjukkan bahwa model klasik masih terbatas dalam memahami konteks pada teks informal. Penelitian selanjutnya disarankan menggunakan metode deep learning seperti BERT dan pelabelan manual untuk hasil lebih akurat.

Pendahuluan

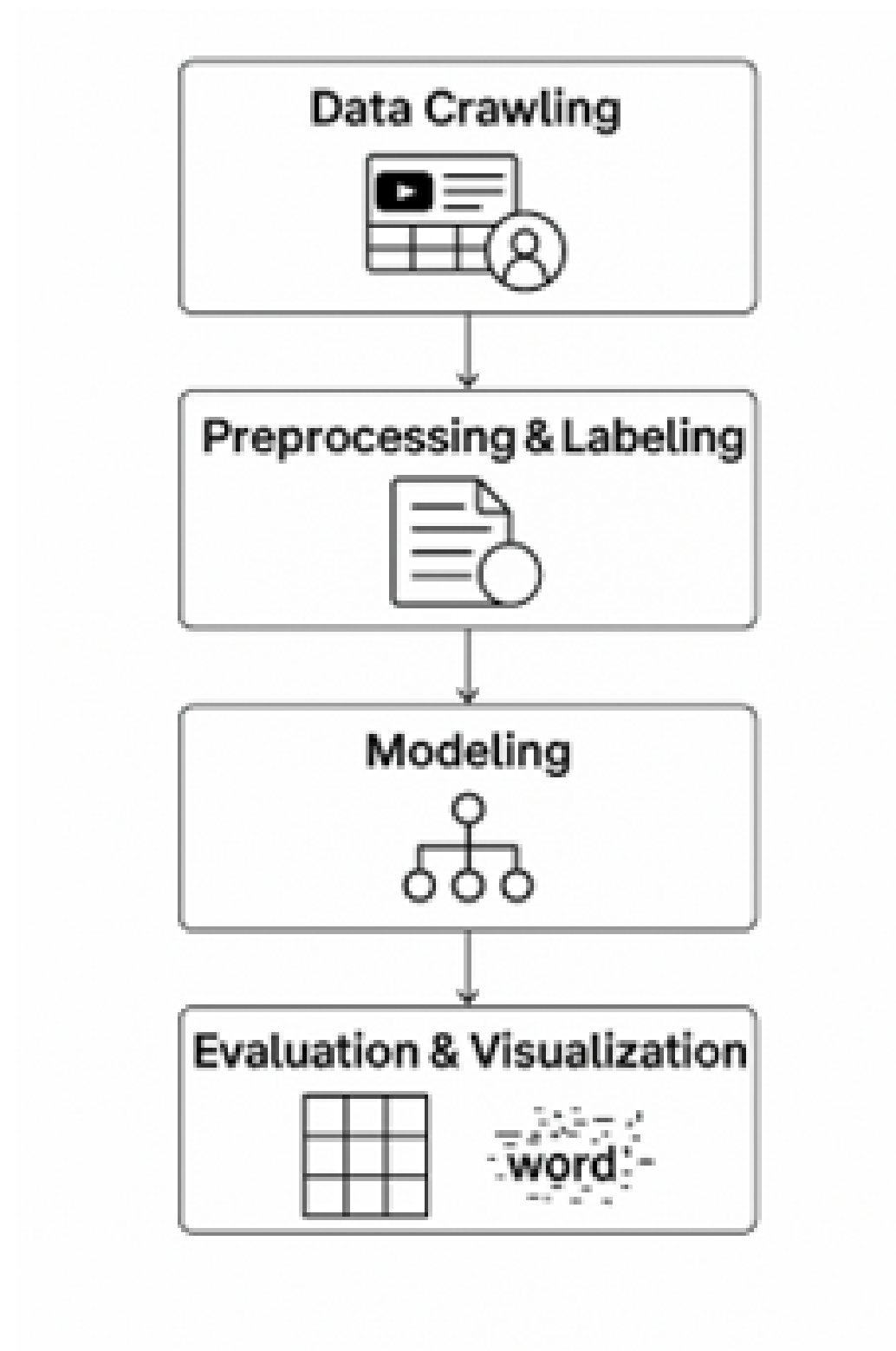
YouTube kini menjadi ruang penting bagi masyarakat untuk menyampaikan opini, termasuk terkait isu politik. Video wawancara Presiden Prabowo Subianto di kanal Najwa Shihab (7 April 2025) menarik perhatian besar dengan lebih dari 2,6 juta penayangan dan ribuan komentar. Komentar tersebut berisi beragam opini, mulai dari dukungan hingga kritik terhadap kebijakan dan kinerja Presiden.

Volume komentar yang sangat besar membuat analisis manual tidak efektif sehingga diperlukan analisis sentimen untuk mengelompokkan opini menjadi positif, negatif, atau netral. Penelitian sebelumnya banyak menggunakan algoritma Naïve Bayes, Random Forest, dan SVM, namun kajian terkait sentimen komentar politik di YouTube masih terbatas.

Orange Data Mining dipilih karena menyediakan workflow visual yang mempermudah preprocessing, pelabelan, dan pelatihan model tanpa perlu pemrograman kompleks. Penelitian ini berkontribusi dalam memperkaya studi analisis sentimen politik di YouTube serta menunjukkan efektivitas penggunaan Orange Data Mining untuk memproses data besar secara efisien.

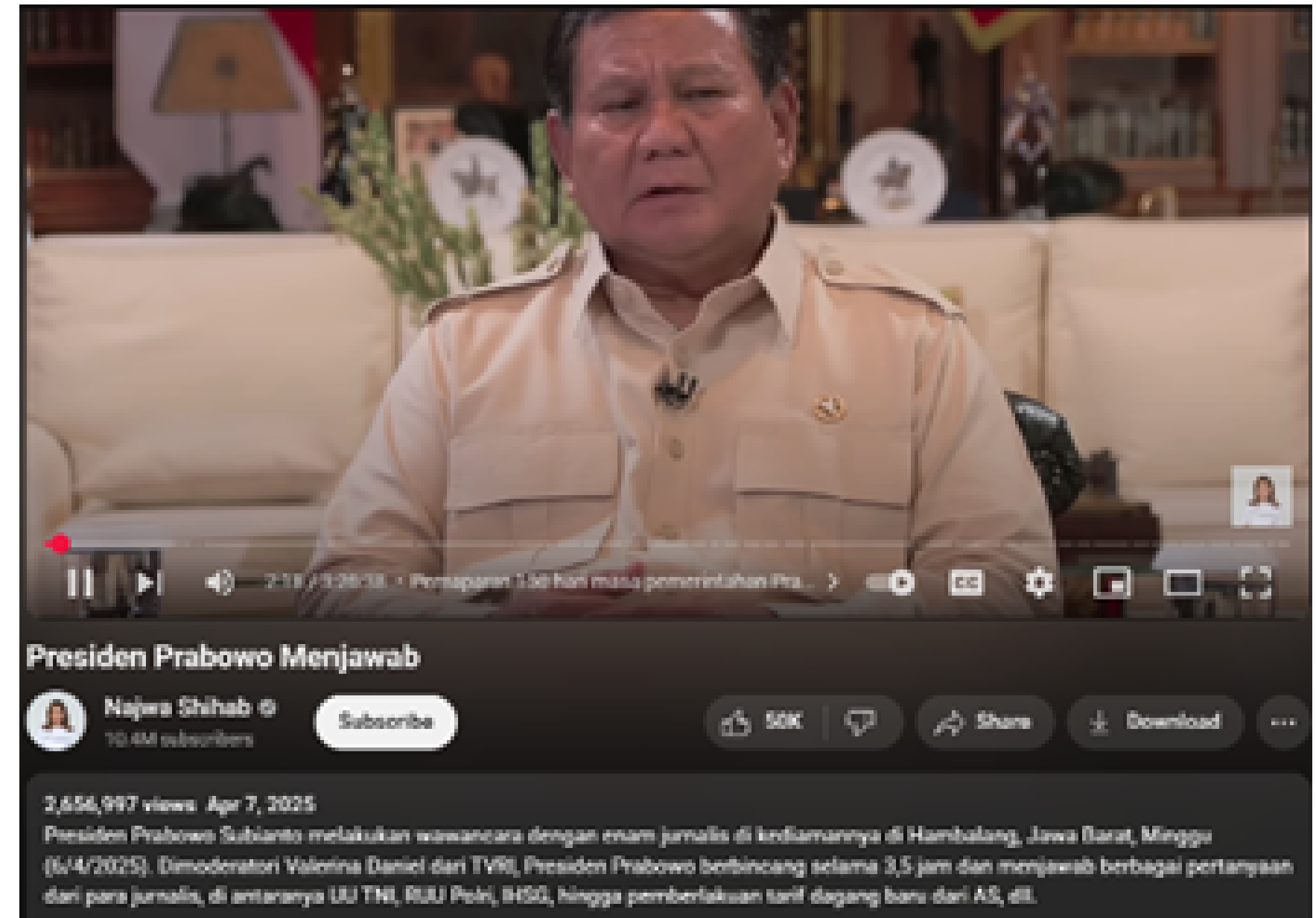


Metode Penelitian



Metode Penelitian

Metode penelitian ini mencakup beberapa tahapan utama. Data dikumpulkan dari komentar video wawancara Presiden Prabowo di YouTube menggunakan YouTube Data API, menghasilkan lebih dari 10.000 komentar sebelum akhirnya disaring menjadi 2.080 komentar yang bersih dan relevan. Selanjutnya, dilakukan preprocessing yang meliputi case folding, cleaning, tokenisasi, stopword removal, dan stemming untuk menghasilkan teks yang siap dianalisis. Pelabelan sentimen dilakukan secara otomatis melalui Sentiment Analysis Module di Orange Data Mining dengan tiga label utama, yaitu positif, negatif, dan netral, disertai validasi manual terhadap sebagian data. Setelah itu, ekstraksi fitur dilakukan menggunakan metode TF-IDF untuk mengubah teks menjadi representasi numerik. Proses pelatihan model menggunakan tiga algoritma—Naïve Bayes, Random Forest, dan Support Vector Machine—dengan evaluasi performa berdasarkan Confusion Matrix, akurasi, presisi, recall, dan F1-score. Tahap akhir melibatkan visualisasi data menggunakan word cloud dan confusion matrix untuk memahami pola dan hasil klasifikasi secara lebih jelas.



Hasil dan Pembahasan

Data Table (1) - Orange

	sentiment_label	comment	sentiment
1	negative	Br sanggup melihat video ini, setelahnya bnr2 merasa yakin melihat bahwa Indonesia memang gelap. Eh maaf br sadar bahwa sy lihatny...	-2.63158
2	positive	Rakyat Bener 2 nggak Salah Pilih Punya Presiden Seperti PK Prabowo Tegas Berwibawa Dan Benar 2 Berpikir Utuk Rakyat.	5
3	neutral	Omon omon	0
4	neutral	Jurnalis nya lebih berwibawa dibanding yang duduk di gedung DPR SUMPAH	0
5	positive	Lanjutkan acara diskusi spt ini kami bangga dgn keterbukaan Bpk PS 🙌🙌🙌	8.33333
6	negative	Bantulah pak jokowi hentikan masalah inazah palsu , kenapa pak	-20
7	neutral	jujur ga relate Pak, maaf Bapak cerita negara dan kepemimpinan mana yaa? hmm	0
8	neutral	baru menit 2 aja dh males	0
9	neutral	🤔🤔🤔	0
10	negative	Di samarinda lagi viral bandar narkoba ditembak mati sesama bandar.apa kinerja presiden.narkoba di smd lagi marak.	-4.7619
11	negative	Hukum itu harus tegas / kejam .biar jera.	-10
12	positive	Mantap pk prabowo pidatonya maju terus pk presiden	12.5
13	positive	Istiqomah mendukung semua kerja Bapak Presiden : "Semoga Bapak Presiden Tetap Sehat, Tetap Happy & Tetap Enjoy". Aamiin.	27.2727
14	positive	Petani sukses bi la pupuk I angsung padaangsu angsung Bu 11:17 11:17 11:17	5.26316

Hasil Pelabelan menggunakan Orange

Dataset penelitian berasal dari ribuan komentar pada video wawancara Presiden Prabowo di YouTube. Data dikumpulkan menggunakan YouTube API dan diperkaya dengan atribut seperti author, published_at, dan title untuk mendukung validasi dan analisis lanjutan. Pelabelan sentimen dilakukan otomatis di Orange Data Mining, yang mengonversi skor numerik menjadi tiga kategori: positif, negatif, dan netral. Sebelum pemodelan, komentar melalui tahapan preprocessing—case folding, cleaning, tokenisasi, stopwords removal, dan stemming—untuk menghasilkan teks bersih yang siap diekstraksi fiturnya dengan TF-IDF.

Preprocessing Data

	full_text	cleaning	stemmed_text	tokenized_text	filtered_text
0	br sanggup melihat video ini, setelahnya bnr2 ...	br sanggup melihat video ini setelahnya bnr2 m...	br sanggup lihat video ini telah bnr2 rasa yak...	[br, sanggup, melihat, video, ini, setelahnya,...	[br, sanggup, melihat, video, setelahnya, bnr2...
1	rakyat bener 2 nggak salah pilih punya preside...	rakyat bener 2 nggak salah pilih punya preside...	rakyat bener 2 nggak salah pilih punya preside...	[rakyat, bener, 2, nggak, salah, pilih, punya,...	[rakyat, bener, 2, nggak, salah, pilih, punya,...
2	omon omon	omon omon	omon omon	[omon, omon]	[omon, omon]
3	jurnalis nya lebih berwibawa dibanding yang du...	jurnalis nya lebih berwibawa dibanding yang du...	jurnalis nya lebih wibawa dibanding yang duduk d...	[jurnalis, nya, lebih, berwibawa, dibanding, y...	[jurnalis, nya, lebih, berwibawa, dibanding, d...
4	lanjutkan acara diskusi spt ini kami bangga dg...	lanjutkan acara diskusi spt ini kami bangga dg...	lanjut acara diskusi spt ini kami bangga dgn k...	[lanjutkan, acara, diskusi, spt, ini, kami, ba...	[lanjutkan, acara, diskusi, spt, kami, bangga,...
5	bantulah pak jokowi hentikan masalah inazah ...	bantulah pak jokowi hentikan masalah inazah pa...	bantu pak jokowi henti masalah inazah palsu ke...	[bantulah, pak, jokowi, hentikan, masalah, ina...	[bantulah, pak, jokowi, hentikan, masalah, ina...
6	jujur ga relate pak, maaf bapak cerita negara ...	jujur ga relate pak maaf bapak cerita negara d...	jujur ga relate pak maaf bapak cerita negara d...	[jujur, ga, relate, pak, maaf, bapak, cerita, ...	[jujur, ga, relate, pak, maaf, bapak, cerita, ...
7	baru menit 2 aja dh males	baru menit 2 aja dh males	baru menit 2 aja dh males	[baru, menit, 2, aja, dh, males]	[baru, menit, 2, aja, dh, males]
8	🤔🤔🤔			[]	[]
9	di samarinda lagi viral bandar narkoba ditemba...	di samarinda lagi viral bandar narkoba ditemba...	di samarinda lagi viral bandar narkoba tembak ...	[di, samarinda, lagi, viral, bandar, narkoba, ...	[samarinda, lagi, viral, bandar, narkoba, dite...

Proses preprocessing dilakukan untuk membersihkan data dari elemen yang tidak relevan agar siap digunakan dalam pelatihan model. Tahapan yang diterapkan meliputi case folding, pembersihan karakter, tokenisasi, stopwords removal, dan stemming. Proses ini bertujuan menghasilkan teks yang seragam dan bebas noise sehingga dapat dikonversi menjadi fitur numerik menggunakan TF-IDF. Pada praktiknya, data awal dibersihkan melalui beberapa tahap—cleaning, stemming, tokenisasi, dan penghapusan stopwords—hingga diperoleh teks yang benar-benar relevan untuk proses klasifikasi.

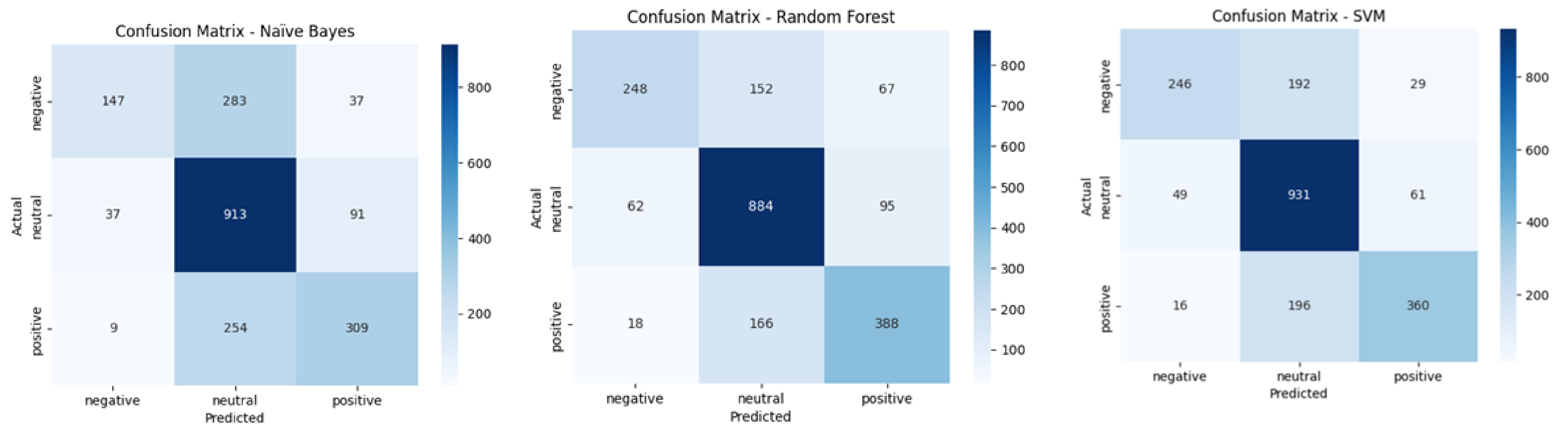
Pemodelan

Tabel 1. Matrix Evaluasi

Model	Accuracy	Precision (avg)	Recall (avg)	F1-score (avg)
Naïve Bayes	66%	0.68 (weighted)	0.66	0.64
SVM	74%	0.75 (weighted)	0.74	0.73
Random Forest	73%	0.73 (weighted)	0.73	0.72

Tahap pemodelan menggunakan tiga algoritma machine learning klasik—Naïve Bayes, SVM, dan Random Forest—untuk membandingkan kinerja dalam klasifikasi sentimen pada 2.080 komentar YouTube yang telah dipreproses dan diberi label. Evaluasi dilakukan menggunakan akurasi, precision, recall, dan f1-score. Naïve Bayes memperoleh akurasi 66% dan f1-score 0,64, dengan kecenderungan mengklasifikasikan banyak komentar sebagai netral. SVM memberikan performa terbaik dengan akurasi 74% dan f1-score 0,73, menunjukkan keseimbangan precision-recall yang baik pada semua kelas. Random Forest menyusul dengan akurasi 73% dan f1-score 0,72, serta stabil dalam mengenali ketiga kelas sentimen. Hasil ini menunjukkan bahwa meskipun model klasik cukup efektif, pendekatan yang lebih canggih seperti BERT atau LSTM berpotensi meningkatkan akurasi karena mampu memahami konteks bahasa secara lebih mendalam.

Visualisasi Hasil

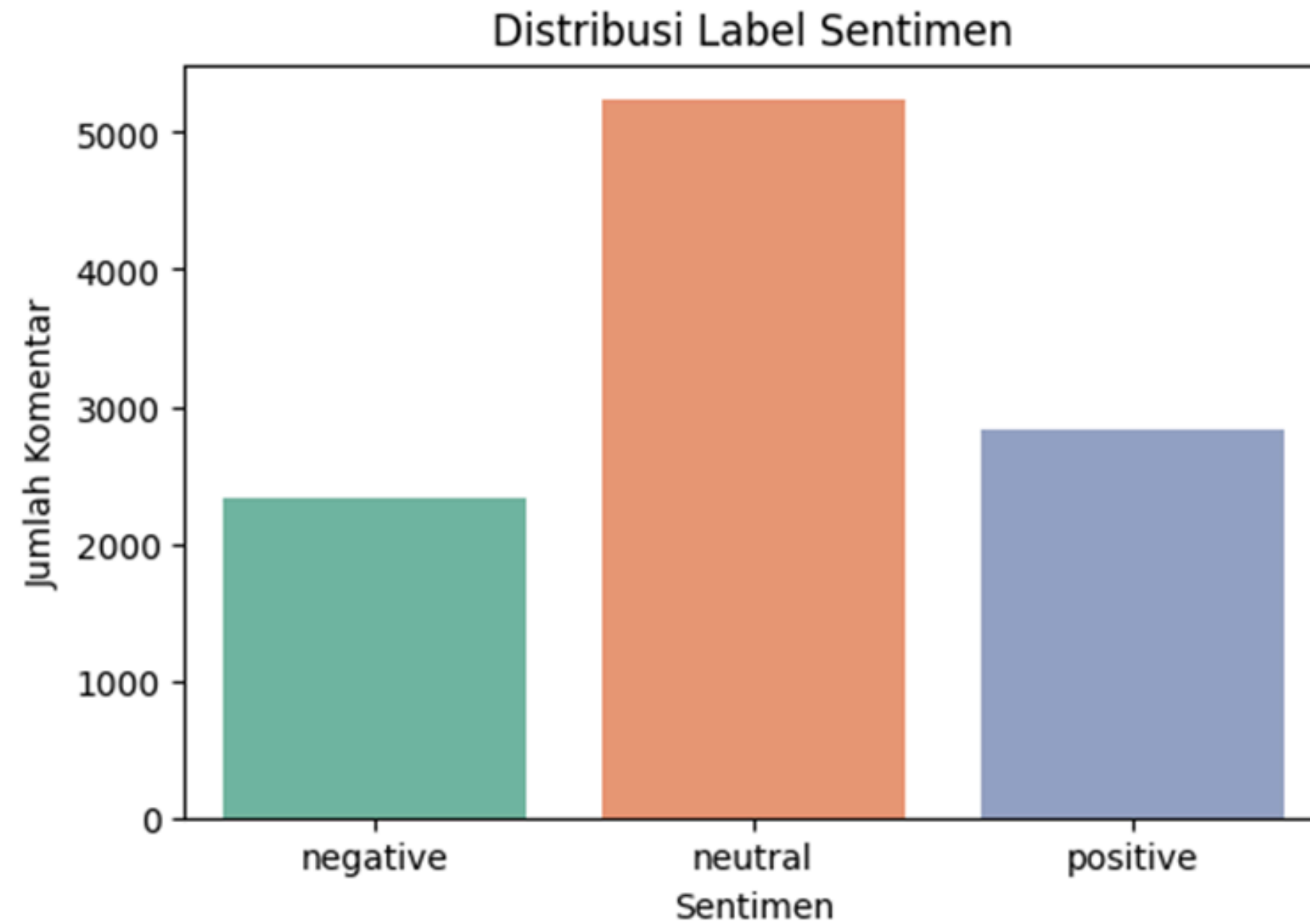


Word Cloud

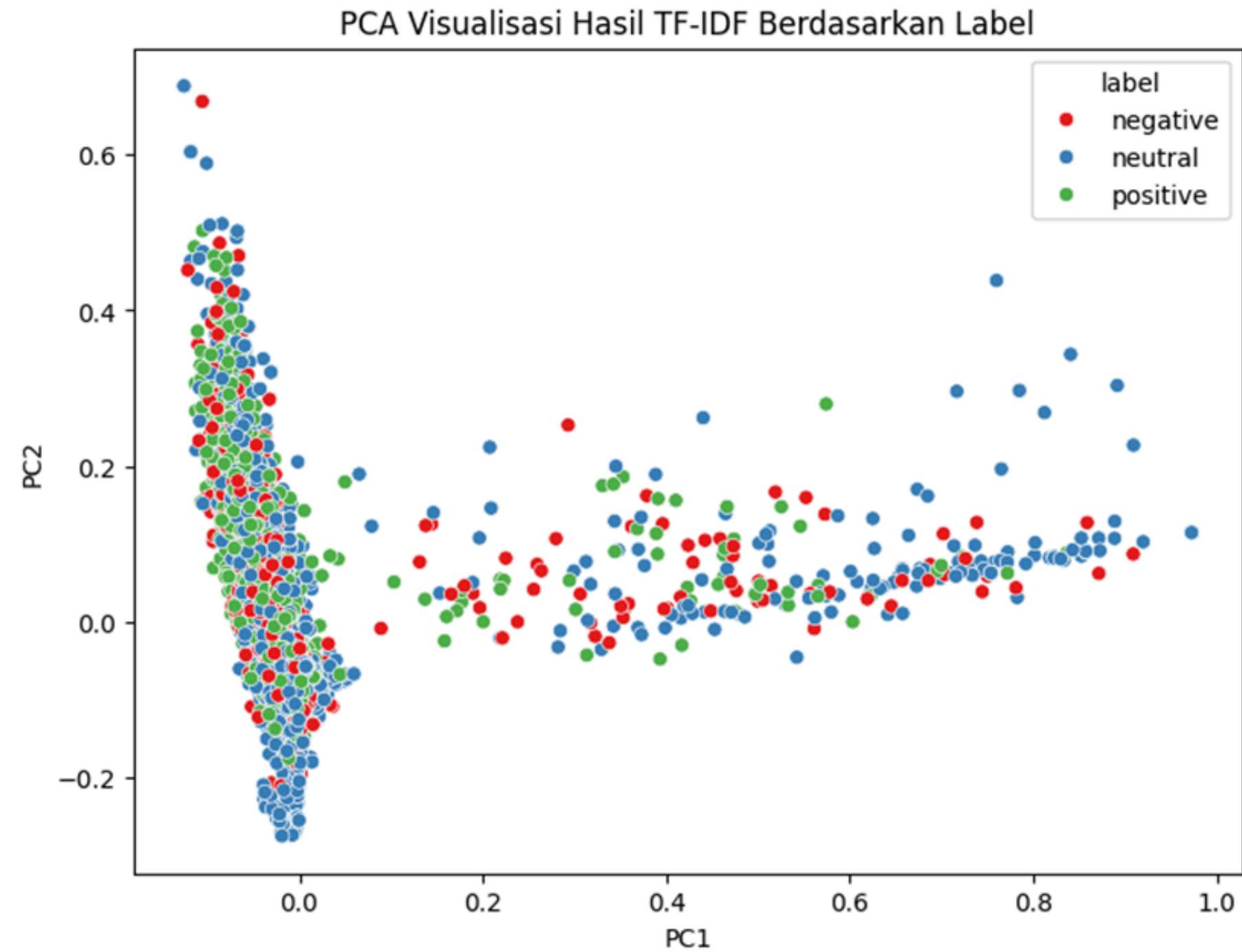
Word Cloud Setelah Preprocessing



Distribusi Label Sentimen



Distribusi Label Sentimen



Kesimpulan

Penelitian ini berhasil menganalisis sentimen komentar publik pada video wawancara Presiden Prabowo di YouTube menggunakan tiga algoritma machine learning. SVM memberikan hasil terbaik dengan akurasi 74%, disusul Random Forest 73% dan Naïve Bayes 66%. Komentar netral mendominasi, sementara komentar negatif sering salah klasifikasi, menunjukkan keterbatasan model klasik dalam memahami konteks bahasa media sosial. Tantangan utama meliputi ketidakseimbangan data dan pelabelan otomatis yang belum sepenuhnya akurat. Untuk penelitian selanjutnya, pelabelan manual sebagian data serta penggunaan model deep learning seperti BERT direkomendasikan guna meningkatkan kualitas klasifikasi. Penelitian ini memberikan gambaran awal mengenai opini publik digital dalam konteks politik di Indonesia.

THANK
YOU

02 May, 2024