

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

<sup>1)</sup> Kristina E.V. Sihombing, <sup>2)</sup>Magdalena A. Ineke Pakereng

<sup>1,2)</sup>Universitas Kristen Satya Wacana, Salatiga

<sup>1)</sup>672022337@student.uksw.edu

---

### INFO ARTIKEL

#### Riwayat Artikel :

Diterima : 17 September 2025

Disetujui : 6 Oktober 2025

---

#### Kata Kunci :

Analisis sentimen, komentar publik, *Machine learning*, *Orange Data Mining*, Prabowo Subianto

---

### ABSTRAK

Media sosial seperti Youtube telah menjadi platform utama dalam menyampaikan opini publik, termasuk dalam isu-isu politik nasional. Penelitian ini bertujuan untuk menganalisis sentimen komentar publik terhadap video wawancara Presiden Prabowo Subianto yang ditayangkan di kanal *Youtube* *Najwa Shihab*. Komentar yang dikumpulkan melalui *Youtube API* diproses menggunakan tahapan *text preprocessing* dan pelabelan *semi-supervised* dengan *Orange Data Mining*. Tiga algoritma *machine learning* *Naïve Bayes*, *Random Forest*, dan *Support Vector Machine* (SVM) digunakan untuk melatih model klasifikasi sentimen. Hasil evaluasi menunjukkan bahwa SVM menghasilkan kinerja terbaik dengan akurasi sebesar 74% dan f1-score tertimbang 0.73. *Random Forest* mencapai akurasi 73%, sedangkan *Naïve Bayes* 66%. Temuan ini menunjukkan bahwa model klasik memiliki keterbatasan dalam memahami konteks dan sentimen tersirat, khususnya pada komentar berbahasa tidak formal. Oleh karena itu, penelitian selanjutnya disarankan menggunakan pendekatan deep learning seperti BERT dan metode pelabelan manual untuk meningkatkan kualitas klasifikasi.

---

### ARTICLE INFO

#### Article History :

Received : Sept 17, 2025

Accepted : Oct 6, 2025

---

#### Keywords:

*machine learning*, *Orange Data Mining*, *Prabowo Subianto*, *public comments*, *Sentiment Analysis*

---

### ABSTRACT

*Social media platforms such as Youtube have become key spaces for the public to express opinions, particularly regarding political issues. This study aims to analyze the sentiment of public comments on a video interview with President Prabowo Subianto broadcasted on *Najwa Shihab's* *Youtube* channel. Comments collected via the *Youtube API* were processed through *text preprocessing* and *semi-supervised labeling* using *Orange Data Mining*. Three *machine learning* algorithms *Naïve Bayes*, *Random Forest*, and *Support Vector Machine* (SVM) were used to train the sentiment classification model. The evaluation results show that SVM achieved the best performance with an accuracy of 74% and a weighted f1-score of 0.73, followed by Random Forest (73%) and *Naïve Bayes* (66%). These findings suggest that classical models face limitations in capturing context and implicit sentiment, especially in informal language comments. Therefore, future research should consider adopting deep learning approaches such as BERT and incorporating manual data labeling to improve classification accuracy.*

---

## 1. PENDAHULUAN

Media sosial telah berkembang menjadi sarana utama bagi masyarakat untuk menyuarakan opini terhadap isu-isu publik. Platform seperti *Youtube*, yang memiliki jutaan pengguna aktif setiap harinya, tidak hanya berfungsi sebagai sumber hiburan dan informasi, tetapi juga menjadi arena diskusi politik dan sosial. Salah satu contoh signifikan dari fenomena ini adalah video wawancara Presiden Prabowo Subianto bersama enam pemimpin redaksi media nasional, yang diunggah di kanal *Youtube* Najwa Shihab pada 7 April 2025. Dalam video berdurasi 3,5 jam tersebut, Presiden Prabowo membahas beragam isu penting mulai dari RUU TNI dan Polri, pasar modal, stunting, hingga proyek strategis Danantara. Video ini memperoleh perhatian besar dari masyarakat dengan lebih dari 2,6 juta penayangan.

Kolom komentar pada video tersebut dipenuhi oleh opini yang beragam dari masyarakat. Sebagian komentar menunjukkan dukungan terhadap kebijakan Presiden, sementara lainnya memberikan kritik terhadap gaya komunikasi, keputusan politik, dan efektivitas pemerintahan selama 150 hari masa kerja. Opini-opini ini mencerminkan persepsi publik yang sangat bernali untuk dianalisis secara sistematis (Hudha, Supriyati and Listyorini, 2022). Namun demikian, volume komentar yang mencapai ribuan entri menjadikan analisis manual tidak efisien dan rentan bias, sehingga diperlukan pendekatan berbasis teknologi seperti analisis sentimen. Untuk itu, diperlukan pendekatan berbasis teknologi, seperti analisis sentimen, yaitu teknik dalam bidang *text mining* yang bertujuan mengklasifikasikan opini dalam bentuk teks menjadi kategori positif, negatif, atau netral.

Analisis sentimen, sebagai salah satu cabang *text mining*, bertujuan mengklasifikasikan opini publik dalam bentuk teks ke dalam kategori positif, negatif, atau netral. Berbagai penelitian sebelumnya telah memanfaatkan algoritma *machine learning* klasik seperti Naive Bayes, Random Forest, dan Support Vector Machine (SVM). Naive Bayes dikenal karena kesederhanaannya dalam menangani teks, Random Forest unggul dalam kestabilan akurasi untuk data kompleks,

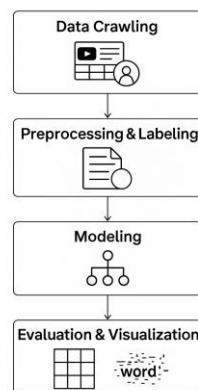
sementara SVM terbukti efektif dalam menangani data berdimensi tinggi. Meski demikian, penelitian yang berfokus pada analisis komentar politik di *YouTube*, khususnya terkait wawancara Presiden, masih relatif terbatas. Sebagian besar studi terdahulu lebih banyak menyoroti sentimen di *Twitter* atau platform berita daring, sehingga terdapat kesenjangan penelitian dalam konteks *YouTube* sebagai ruang diskusi politik.

Selain itu, keterbatasan yang kerap muncul adalah penggunaan alat analisis yang bersifat kompleks dan memerlukan kemampuan pemrograman tingkat lanjut. Dalam hal ini, Orange Data Mining menawarkan keunggulan sebagai platform visual berbasis *workflow* yang mempermudah integrasi *preprocessing*, pelabelan data, pelatihan model, hingga evaluasi performa. Dukungan pada metode semi-supervised learning juga menjadikan Orange relevan digunakan dalam mengolah data komentar yang berjumlah besar dengan keterbatasan pelabelan manual.

Berdasarkan uraian tersebut, penelitian ini memiliki kontribusi dalam dua aspek utama. Pertama, secara teoretis, penelitian ini memperkaya kajian analisis sentimen dengan menyoroti opini publik di *YouTube* terkait isu politik, yang masih jarang dieksplorasi dibandingkan platform lain. Kedua, secara praktis, penelitian ini menunjukkan efektivitas penggunaan Orange Data Mining dalam mempermudah proses analisis sentimen berbasis *machine learning*, sehingga dapat menjadi rujukan bagi peneliti maupun praktisi yang ingin mengevaluasi persepsi publik secara lebih efisien. Dengan demikian, penelitian ini tidak hanya menyajikan gambaran kuantitatif mengenai opini publik terhadap wawancara Presiden Prabowo, tetapi juga menawarkan pendekatan metodologis yang relevan untuk kajian analisis politik digital di Indonesia.

## 2. METODE

Metode yang digunakan dalam penelitian ini mengikuti tahapan dalam proses analisis data teks berbasis *machine learning* yang meliputi: pengumpulan data, *preprocessing*, pelabelan data, pelatihan model, evaluasi performa, dan visualisasi hasil (Fitri, 2020). Perhatikan di Gambar 1.



Gambar 1. Tahapan Penelitian

### a. Pengumpulan Data

Data penelitian diperoleh dari komentar pada video wawancara Presiden Prabowo Subianto yang diunggah di kanal *YouTube* Najwa Shihab pada 7 April 2025, menggunakan *YouTube Data API v3*. Sebanyak lebih dari 10.000 komentar berhasil di-*crawling* dan disimpan dalam format CSV. Dataset tersebut dilengkapi atribut *author*, *published\_at*, dan *title*, sehingga memungkinkan validasi identitas akun serta analisis distribusi waktu komentar (Imran, Karim, & Ningsih, 2024).

Namun, dari total komentar yang terkumpul, hanya 2.080 komentar yang dapat digunakan dalam analisis. Penyusutan data ini terjadi akibat proses pembersihan (cleaning) yang menghapus komentar duplikat, spam, iklan, emotikon kosong, serta komentar yang hanya berisi tautan. Selain itu, komentar yang tidak relevan dengan topik wawancara juga dieliminasi. Dengan demikian, jumlah data akhir merepresentasikan komentar yang lebih bersih dan valid untuk dianalisis lebih lanjut.

Gambar 2 memperlihatkan tampilan video asli tempat komentar dikumpulkan, yang menunjukkan judul video, jumlah tayangan, serta konteks durasi dan tanggal unggah.



Gambar 2. Tampilan Video Youtube Wawancara Presiden Prabowo Subianto

Gambar 3 memperlihatkan tampilan data hasil *crawling* lengkap dengan atribut tambahan dari *Youtube API*.

title	author	comment	published_at
1	@SipirCity	Br sangat suka nonton video ini, sebaliknya br2 juga yakin yakin milih bahwa Indonesia memang gelap. Eh ma...	2025-05-06 23:57:42
2	@panduan_4ly	Rakyat Bener 2 negal Salat Pimpinan Sopir! Pak Prabowo Tegar Berwibawa Dan Bener 2 Bener...	2025-05-06 18:53:59
3	@SASTRALEGGR802	Jurnalist nya lebih berwibawa dibanding yang duluk di gelung DPRP SUMPAH	2025-05-06 18:05:16
4	@CandraLame_kbn	Lanjutkan acara diskusi ikti kimi dengan dgn ktemuan Bpk PS	2025-05-06 16:50:39
5	@ekoproses988	Bantulah pak jokowi... berikan masalah hasil pulu... kerangka pak...	2025-05-06 16:48:29
6	@AnalisaNasya	Bantulah pak jokowi... mudah bagai cinta negara dan kepentingan manusia juga? Hem...	2025-05-06 07:58:41
7	@muhammadrudinwijaya404	Bantulah pak jokowi... mudah bagai cinta negara dan kepentingan manusia juga? Hem...	2025-05-05 21:15:43
8	@RahmatulHuda	Hem...	2025-05-05 16:06:55
9	@RahmatulHuda	Hem...	2025-05-05 16:06:55
10	@Sugiharto_012	Hukum itu harus tetap / Adil. Boleh jadi...	2025-05-05 15:58:03
11	@Sugiharto_012	Menaruh pk prabowo pedatanya maha heros per Presiden...	2025-05-05 14:59:06
12	@Sugiharto_012	Integritas dan pengabdian semakin tinggi Presiden Prabowo	2025-05-05 14:59:06
13	@Baulahmeti_154	Integritas dan pengabdian semakin tinggi Presiden Prabowo	2025-05-05 14:59:06
14	@Mulyadi_1990	Pak, mohon pakah lokak dg dr praktika dr pemerintah barang lauk terhadama benereng Regol + servis & kstad dlm...	2025-05-05 09:53:01
15	@Tenggakpanitia2356	Ahamahamaham temui maha mene orang pak...	2025-05-05 09:53:01
16	@Amirulhisham174	Gpp...	2025-05-05 09:53:01
17	@Amirulhisham174	Gpp...	2025-05-05 09:53:01
18	@Soswono_91F	Gpp...	2025-05-05 09:53:01
19	@Rafiq_imamzak	Sehat selalu Pak presiden yg yg B. Setelah dan sukses pak menjaga Indonesia emas... NIKHIL hargaa mati...	2025-05-04 20:47:37
20	@Tehakimzaki	Presiden gitung...	2025-05-04 18:42:35
21	@Tehakimzaki	Presiden gitung...	2025-05-04 18:42:35

Gambar 3. Tampilan Dataset Komentar *Youtube* hasil *Crawling*

### b. Preprocessing Data

Tahap *Preprocessing Data* dilakukan dengan tahapan *case folding*, *cleaning*, *tokenisasi*, *stopword removal*, dan *stemming* menggunakan *library NLP* Bahasa Indonesia (Muhyat, Fauzi and Indra, 2023). Proses ini menghasilkan data teks yang bersih dan siap untuk diekstraksi fiturnya (Adelia et al., 2023).

### c. Pelabelan Data

Pelabelan sentimen dilakukan dengan memanfaatkan *Sentiment Analysis Module* di Orange Data Mining (Anggraini & Suroyo, 2019). Modul ini menghasilkan skor numerik yang kemudian dikategorikan menjadi tiga kelas sentimen: positif, negatif, dan netral. Pengelompokan dilakukan dengan bantuan *Python Script* di dalam Orange (Saputra & Pribadi, 2023).

Meskipun pelabelan dilakukan secara otomatis, penelitian ini menyadari pentingnya validasi manual untuk menjamin reliabilitas. Oleh karena itu, sebagian data (10% dari total dataset) dilakukan pengecekan manual secara acak untuk membandingkan hasil label otomatis dengan interpretasi peneliti. Hasil pengecekan ini digunakan sebagai dasar validasi keandalan model pelabelan. Langkah ini penting karena analisis sentimen berbasis *machine learning* rentan terhadap bias algoritmik dan ambiguitas bahasa (Hudha, Supriyati, & Listyorini, 2022).



Gambar 4. *Workflow Analisis Sentimen di Orange Data Mining*

Selain itu, Gambar 5 memperlihatkan *Word cloud* hasil dari proses analisis sentimen yang menunjukkan kata-kata yang paling sering muncul dalam komentar publik. Kata “yang”, “pak”, “prabowo”, dan “presiden” menjadi kata dominan yang menggambarkan topik diskusi (Sapanji, Hamdani and Harahap, 2023).



Gambar 5. *Word Cloud* Hasil Analisis Sentimen di *Orange*

#### d. Ekstraksi Fitur

Tahap Ekstraksi Fitur (Muhyat, Fauzi and Indra, 2023) terhadap komentar yang telah melalui *preprocessing* diekstrak fiturnya menggunakan metode TF-IDF yang menghasilkan representasi numerik untuk proses klasifikasi.

### e. Pelatihan Model

Tahap Pelatihan Model yang digunakan dalam penelitian ini meliputi:

1. Naive bayes: efektif untuk klasifikasi teks dengan asumsi independensi antar fitur(Adelia et al., 2023)(Informasi et al., 2024b).
  2. Random Forest: unggul dalam akurasi dan kestabilan(Fitri, 2020)(Syafia, Hidayattullah and Suteddy, 2023)(Artanto et al., 2024).
  3. Support Vector Machine (SVM): cocok untuk data berdimensi tinggi seperti teks yang direpresentasikan dengan TF-IDF (Syafia, Hidayattullah and Suteddy,

2023). Pelatihan dilakukan di Google colaboratory dengan dataset hasil pelabelan dan preprocessing (Imran, Karim and Ningsih, 2024).

## f. Evaluasi Model

Evaluasi performa model dilakukan dengan *Confusion Matrix* serta metrik akurasi, presisi, recall, dan F1-score, yang dirumuskan sebagai berikut: *Confusion Matrix* (Imran, Karim and Ningsih, 2024) (Hudha, Supriyati and Listyorini, 2022)

- a. Akurasi :

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

- b. *Precision:*

$$a. \quad Precision = \frac{TP}{TP+FP}$$

- c. *Recall*:

$$d. \ Recall = \frac{TP}{TP+FN}$$

- e. *F1-Score* :

$$a. F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Model terbaik dipilih berdasarkan keseimbangan antara akurasi dan *F1-score*.

## g. Visualisasi Hasil

Tahap visualisasi dilakukan menggunakan *Word cloud* untuk menunjukkan kata-kata dominan dan *Confusion Matrix* untuk menampilkan hasil klasifikasi setiap model secara visual. Tahapan-tahapan ini dirancang untuk memaksimalkan efektivitas dalam menangkap opini publik melalui teks digital yang masif dan tidak terstruktur. Dengan metodologi ini, penelitian berupaya memastikan bahwa meskipun pelabelan awal dilakukan secara otomatis, tetap ada upaya validasi reliabilitas untuk memperkuat keabsahan hasil analisis.

### 3. HASIL DAN PEMBAHASAN

Dataset yang dianalisis terdiri dari 2.080 komentar publik yang dikumpulkan dari video wawancara Presiden Prabowo Subianto di kanal *YouTube* *Najwa Shihab*. Komentar ini

mencerminkan beragam opini masyarakat, mulai dari dukungan terhadap kebijakan pemerintah hingga kritik terhadap gaya komunikasi Presiden maupun efektivitas pemerintahan. komentar yang diambil dari video wawancara Presiden Prabowo Subianto di kanal *Youtube* Najwa Shihab. Komentar-komentar tersebut mencerminkan opini publik yang beragam, mulai dari pujian terhadap kebijakan pemerintah hingga kritik tajam terhadap gaya komunikasi dan efektivitas kinerja pemerintahan. Data diperoleh melalui *Youtube API*, dilengkapi dengan atribut tambahan seperti nama pengguna (*author*), tanggal publikasi komentar (*published\_at*), dan judul video (*title*), sesuai praktik pada penelitian sebelumnya. Atribut tambahan ini berfungsi untuk validasi dan memungkinkan perluasan analisis, seperti pengamatan temporal dan segmentasi akun. Format data disimpan dalam CSV, yang mempermudah proses integrasi dengan platform *Orange* dan *Google colab*.

Proses pelabelan Sentimen dilakukan sepenuhnya secara otomatis di *Orange Data Mining*. Modul 'Sentiment Analysis' dalam *Orange* memberikan skor numerik terhadap setiap komentar berdasarkan analisis teks. Skor ini kemudian dikategorikan menjadi tiga kelas sentimen: positif, negatif, dan netral, menggunakan *script Python* tambahan dalam *Orange*. Gambar 4 menunjukkan visualisasi alur kerja (*workflow*) pada *Orange* yang digunakan untuk proses pelabelan sentimen (Bayu and Wurijanto, 2024) (Jamil and Hanafi, 2025) (Informasi et al., 2024a).



Gambar 6. Hasil Pelabelan oleh *Orange*

Pelabelan komentar pada Gambar 6, dilakukan otomatis menggunakan *Orange*. Skor sentimen dikonversi menjadi label sentimen menggunakan ambang batas, yaitu: skor  $> 0$  sebagai positif, skor  $< 0$  sebagai negatif, dan skor  $= 0$  sebagai netral. Hasil pelabelan ini kemudian digunakan sebagai *dataset* untuk proses pelatihan model di *Google colab*.

Preprocessing dilakukan dengan tahapan *case folding*, *cleaning*, *tokenization*, *stopword removal*, dan *stemming*. Hasil preprocessing menunjukkan dominasi kata-kata seperti *pk*, *prabowo*, *rakyat*, dan *negara*, yang menandakan diskursus publik lebih banyak berfokus pada figur kepemimpinan serta kebijakan negara. Sementara itu, kata informal seperti *gak*, *omong*, dan *bapak* menandakan gaya komunikasi santai khas ruang publik digital, yang berbeda dengan diskursus resmi di media arus utama.

Pada *Preprocessing data* yang telah dikumpulkan. Proses *preprocessing* dilakukan untuk menghilangkan elemen-elemen yang tidak relevan dan menyiapkan data agar dapat digunakan dalam pelatihan model. Langkah-langkah *preprocessing* meliputi *case folding*, pembersihan karakter, tokenisasi, *stopword removal*, dan *stemming* (Imran, Karim and Ningsih, 2024) (Burnama, Rosid and Azizah, n.d.). Tahapan ini bertujuan menghasilkan korpus teks bersih dan seragam agar dapat dikonversi menjadi fitur numerik menggunakan metode seperti *TF-IDF* (Winarno, Kurniabudi and Dr. Hendri, S.Kom., S.H., M.S.I., M.H, 2023) (Informasi et al., 2024b). Proses ini penting karena model *machine learning* sensitif terhadap *noise* dalam teks (Fitri, 2020).



Gambar 7. Hasil Preprocessing

Gambar 7. menunjukkan tahapan hasil *preprocessing* yang dilakukan di *Google colab*.

Data awal (*full\_text*) dibersihkan melalui beberapa tahap: pembersihan karakter (*cleaning*), stemming (*stemmed\_text*), tokenisasi (*tokenized\_text*), dan penghapusan *stopword* (*filtered\_text*). Tahapan ini memastikan data input untuk model dalam bentuk teks bersih dan relevan untuk proses klasifikasi.

Tahap pemodelan pada penelitian ini menggunakan tiga algoritma *machine learning* klasik untuk membandingkan kinerja dalam tugas klasifikasi sentimen, yaitu *Naive Bayes*, *Support Vector Machine (SVM)*, dan *Random*

*Forest.* Dataset yang digunakan berjumlah 2.080 komentar Youtube yang telah melalui tahapan *preprocessing* dan pelabelan. Setiap model dievaluasi berdasarkan metrik *precision*, *recall*, *f1-score*, serta akurasi keseluruhan.

a. *Naive Bayes*

Model *Naive Bayes* menunjukkan akurasi sebesar 66%. Metrik *precision* tertinggi berada pada kelas negatif (0,76), namun *recall* yang dihasilkan rendah (0,31), mengindikasikan bahwa model cenderung hanya mengenali komentar negatif jika sangat yakin. Sementara itu, kelas netral memiliki *recall* yang sangat tinggi (0,88) namun *precision* rendah (0,63), yang menunjukkan kecenderungan model untuk menggeneralisasi komentar sebagai netral. *F1-score* rata-rata tertimbang (*weighted avg*) tercatat 0,64. Hal ini menggambarkan keterbatasan algoritma sederhana dalam menangkap nuansa bahasa yang sarkastis atau implisit, yang sering dipakai netizen saat menyinggung isu politik. Temuan ini sejalan dengan studi sebelumnya yang menunjukkan bahwa *Naive Bayes* bekerja cepat tetapi kurang akurat dalam menangani data sosial media yang kaya konteks.

b. *Support Vector Machine (SVM)*

Model *SVM* memperoleh akurasi tertinggi yaitu 74%, dengan *f1-score* tertimbang sebesar 0,73. Model ini menunjukkan keseimbangan yang lebih baik dalam *precision* dan *recall* di ketiga kelas. Kelas netral memperoleh *f1-score* tertinggi (0,79), sedangkan kelas negatif dan positif masing-masing memperoleh 0,63 dan 0,70. Keunggulan *SVM* dalam menangani data berdimensi tinggi seperti vektor *TF-IDF* telah dibuktikan pula dalam studi sebelumnya. Keunggulan ini mengonfirmasi efektivitas *SVM* pada data berdimensi tinggi seperti *TF-IDF* (Syafia, Hidayattullah, & Suteddy, 2023).

c. *Random Forest*

*Random Forest* juga menunjukkan performa yang kompetitif dengan akurasi 73%, dan *f1-score* tertimbang sebesar 0,72. Meskipun sedikit lebih rendah dibanding *SVM*, model ini menunjukkan keseimbangan performa pada ketiga kelas. Kelas netral tetap memperoleh skor tertinggi (*f1-score* 0,79),

diikuti kelas positif (0,69) dan negatif (0,62). *Random Forest* sebelumnya juga dilaporkan memiliki keunggulan dalam stabilitas model untuk data sosial media yang kompleks. Kinerja stabil menunjukkan bahwa algoritma ini mampu menangkap pola kompleks, meskipun tetap menghadapi bias terhadap kelas mayoritas (netral).

Tabel 1. merangkum metrik evaluasi utama dari ketiga model.

Tabel 1. *Matrix Evaluasi*

<i>Model</i>	<i>Accurac y</i>	<i>Precisio n (avg)</i>	<i>Recal l (avg)</i>	<i>F1- scor e (avg )</i>
<i>Naive Bayes</i>	66%	0.68 ( <i>weighte d</i> )	0.66	0.64
<i>SVM</i>	74%	0.75 ( <i>weighte d</i> )	0.74	0.73
<i>Random Forest</i>	73%	0.73 ( <i>weighte d</i> )	0.73	0.72

Hasil ini memperkuat temuan bahwa meskipun algoritma klasik masih efektif, kompleksitas data komentar publik di media sosial mengharuskan model yang mampu menangani konteks bahasa yang lebih dalam. Oleh karena itu, pengembangan model berbasis *Deep Learning* seperti *BERT* atau *LSTM* menjadi opsi yang layak untuk eksplorasi di masa depan (Jamil and Hanafi, 2025).

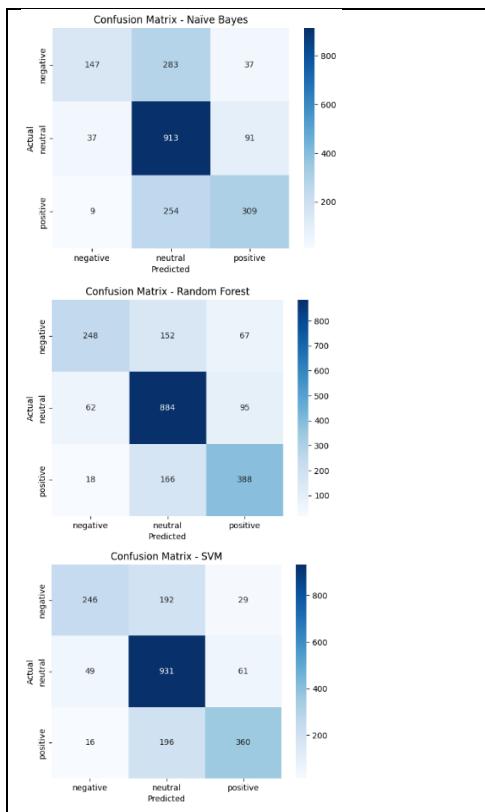
Visualisasi hasil dilakukan untuk memperkuat pemahaman terhadap data dan performa model secara eksploratif. Beberapa jenis visualisasi digunakan dalam penelitian ini, yaitu *confusion matrix*, *word cloud*, distribusi label, dan reduksi dimensi *PCA* terhadap fitur *TF-IDF* (Pahtonni and Jati, 2024).

### **CONFUSION MATRIX**

Gambar 8 menampilkan *confusion matrix* dari ketiga algoritma yang digunakan.

a. *Naive Bayes* menunjukkan banyak kesalahan dalam mengklasifikasikan komentar negatif sebagai netral (283 kasus) dan komentar positif sebagai netral (254 kasus). Hal ini

- menyebabkan *recall* kelas negatif rendah (0.31).
- b. *Random Forest* menunjukkan peningkatan signifikan, dengan jumlah prediksi benar meningkat untuk semua kelas. Misalnya, 248 komentar negatif berhasil dikenali dengan benar dibandingkan hanya 147 pada *Naive Bayes*.
- c. *SVM* menghasilkan distribusi prediksi paling seimbang, dengan nilai diagonal (prediksi benar) tertinggi di semua kelas, terutama kelas netral dan positif. Hal ini mendukung hasil evaluasi sebelumnya di mana *SVM* memiliki *f1-score* terbaik di antara model yang diuji.



Gambar 8. *Confusion Matriks* Dengan Masing-Masing Model

### WORLD CLOUD

Gambar 9 menunjukkan *word cloud* hasil *preprocessing*. Kata-kata seperti “*pak*”, “*prabowo*”, “*rakyat*”, “*presiden*”, dan “*negara*” mendominasi komentar, mencerminkan bahwa topik diskusi publik terfokus pada sosok tokoh, institusi, dan kebijakan nasional. Kata-kata informal seperti “*gak*”, “*aja*”, “*bapak*”, dan “*omong*” menunjukkan gaya bahasa santai yang khas

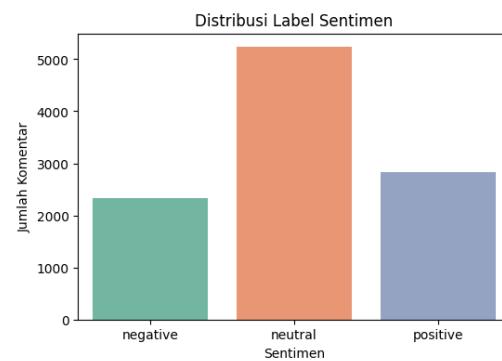
dalam komentar *Youtube*. *Word cloud* seperti ini umum digunakan dalam analisis sentimen awal untuk mengenali kata dominan.



Gambar 9. *Word cloud* Setelah *Preprocessing*

### DISTRIBUSI LABEL SENTIMEN

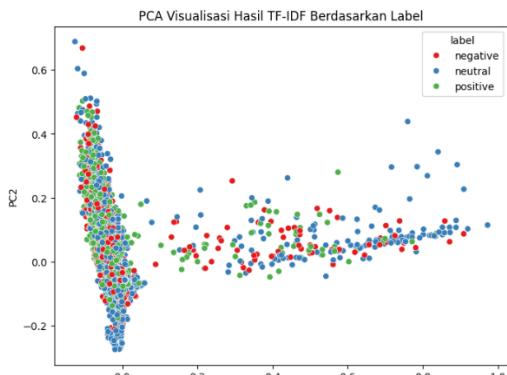
Gambar 10 memperlihatkan distribusi label setelah proses pelabelan sentimen. Kelas netral mendominasi dengan lebih dari 5.000 komentar, disusul oleh kelas positif dan negatif. Ketidakseimbangan ini menjadi salah satu tantangan dalam pelatihan model, karena model cenderung bias terhadap kelas mayoritas.



Gambar 10. Distribusi Label Sentimen

### VISUALISASI PCA

Gambar 11 memperlihatkan hasil reduksi dimensi menggunakan *PCA* (*Principal Component Analysis*) terhadap vektor *TF-IDF* (Faiq, Putro and Setiawan, 2022). Titik-titik berwarna menunjukkan label sentimen yang terdiri dari merah untuk negatif, biru untuk netral, dan hijau untuk positif. Sebagian besar titik cenderung terkonsentrasi, menunjukkan bahwa pemisahan antar kelas tidak terlalu jelas di ruang dua dimensi. Hal ini juga menunjukkan keterbatasan *TF-IDF* dalam menangkap nuansa semantik dalam teks



Gambar 11. PCA

### Analisis Konteks Sosial-Politik

Distribusi sentimen menunjukkan bahwa mayoritas komentar bersifat netral, dengan jumlah positif dan negatif relatif seimbang. Dominasi sentimen netral dapat dipahami sebagai bentuk “observasi” atau *wait and see* publik terhadap pemerintahan Prabowo yang masih berada dalam 150 hari pertama masa kerja. Hal ini sejalan dengan tradisi politik Indonesia, di mana masyarakat cenderung menilai kepemimpinan baru secara bertahap sebelum bersikap lebih kritis.

Komentar positif umumnya muncul terkait harapan pada kebijakan strategis, seperti isu ketahanan pangan dan program stunting. Sebaliknya, komentar negatif banyak berkaitan dengan gaya komunikasi Presiden yang dianggap kurang lugas, serta skeptisme terhadap realisasi proyek besar seperti Danantara. Fenomena ini menegaskan bahwa media sosial bukan sekadar ruang hiburan, tetapi juga arena politik informal di mana warga menegosiasikan pandangan mereka terhadap kekuasaan.

### Karakter Bahasa Komentar

Bahasa komentar mencerminkan ciri khas diskursus publik di media sosial Indonesia. Banyak komentar menggunakan bahasa sehari-hari, humor, atau sindiran, bahkan mencampurkan bahasa daerah dan bahasa Inggris. Bentuk ekspresi ini sering tidak tertangkap oleh analisis otomatis berbasis kosakata standar. Misalnya, komentar seperti “*pak, janji jangan PHP lagi ya*” secara literal tampak netral, tetapi dalam konteks politik bermakna kritik. Hal ini menjelaskan mengapa model cenderung bias pada kelas netral.

Selain itu, penggunaan istilah populer seperti “*strong leader*”, “*gaskeun*”, atau “*makan gaji buta*” menggambarkan perpaduan antara gaya bahasa milenial dan wacana politik formal. Fenomena ini menegaskan bahwa analisis sentimen berbasis algoritma perlu disandingkan dengan pendekatan linguistik atau *discourse analysis* agar makna sosial-politik lebih akurat.

Secara keseluruhan, hasil penelitian menunjukkan bahwa opini publik terhadap Presiden Prabowo di YouTube masih didominasi sentimen netral, dengan perdebatan positif dan negatif yang relatif seimbang. Dari sisi teknis, SVM terbukti paling efektif dalam mengklasifikasikan komentar. Namun, keterbatasan metode tradisional dalam membaca konteks sosial-politik dan bahasa sarkastis mengindikasikan perlunya pengembangan model berbasis *deep learning* (misalnya BERT atau IndoBERT) yang mampu menangkap nuansa semantik lebih dalam. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi teknis dalam pemodelan, tetapi juga memperlihatkan bagaimana media sosial menjadi cermin dinamika politik digital di Indonesia.

## 4. PENUTUP

### Kesimpulan

Penelitian ini berhasil menganalisis sentimen komentar publik pada video wawancara Presiden Prabowo Subianto di YouTube dengan pendekatan *machine learning*. Tiga algoritma yaitu Naive Bayes, Random Forest, dan Support Vector Machine (SVM) diimplementasikan untuk membandingkan performa dalam klasifikasi sentimen. Hasil evaluasi menunjukkan bahwa SVM memberikan kinerja terbaik dengan akurasi sebesar 74% dan *f1-score* tertimbang 0.73, disusul oleh Random Forest (73%) dan Naive Bayes (66%). Visualisasi *word cloud* dan distribusi label menunjukkan dominasi komentar netral, sementara *confusion matrix* mengungkapkan bahwa komentar negatif paling sering salah klasifikasi. Hal ini menegaskan bahwa model tradisional masih memiliki keterbatasan dalam mengenali sentimen eksplisit dan implisit di media sosial (Jamil & Hanafi, 2025).

Implikasi praktis dari penelitian ini adalah hasil analisis sentimen dapat

dimanfaatkan oleh pemerintah, media, dan pengambil kebijakan untuk memahami kecenderungan opini publik secara cepat dan efisien, khususnya dalam isu-isu sosial-politik. Selain itu, model yang dibangun juga dapat diaplikasikan oleh pelaku industri media digital maupun konsultan komunikasi politik untuk melakukan pemetaan opini sebelum mengambil langkah strategis.

Keterbatasan utama penelitian ini meliputi distribusi data yang tidak seimbang, kualitas pelabelan otomatis, serta representasi fitur yang belum mempertimbangkan konteks kalimat. Oleh karena itu, penelitian selanjutnya disarankan untuk: (1) menggunakan metode pelabelan manual sebagian data untuk meningkatkan akurasi, (2) mengadopsi model berbasis *deep learning* seperti BERT yang lebih mampu memahami makna kontekstual, dan (3) memperluas cakupan data dari berbagai platform media sosial untuk mendapatkan gambaran opini publik yang lebih komprehensif. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi teoretis dalam pengembangan analisis sentimen di Indonesia, tetapi juga dapat menjadi dasar dalam perumusan strategi komunikasi publik yang lebih efektif dan responsif.

## 5. DAFTAR PUSTAKA

- Adelia, S., Milanda, E., Santari, J., Kesuma, D.T., Silvia, E. And Kurniawan, F., 2023. Analisis Sentimen Belajar Programming Pada Media Sosial Youtube Menggunakan Algoritma Klasifikasi Naive Bayes. *Journal of Information Technology Ampera*, [online] 4(3), pp.254–264. Available at: <<https://journal-computing.org/index.php/journal-ita/article/view/430>>.
- Adi, S., Mola, S., Iskandar, I.M., Elisa, J., Dimu, P. And Seran, W.Y., 2024. Analisis Sentimen Pembangunan Ibu Kota Negara Indonesia Di Youtube Menggunakan Metode Naïve Bayes , Dan K-Nearest Neighbor. 15, pp.151–157.
- Anggraini, N. And Suroyo, H., 2019. Perbandingan Analisis Sentimen Terhadap Digital Payment “ T-cash dan Go-pay” Di Sosial Media Menggunakan Orange Data Mining. *Journal of Information Systems and Informatics*, [online] 1(1), pp.152–163. Available at: <<http://journal-isi.org/index.php/isi>>.
- Artanto, F.A., Teknik, F., Muhammadiyah, U. And Pekalongan, P., 2024. Implementasi Algoritma Random Forest dan Model Bag of Words Dalam Analisis Sentimen Mengenai E-Materai. 4(2), pp.139–145. <https://doi.org/10.54259/satesi.v4i2.3240>
- Bayu, L.K. And Wurijanto, T., 2024. Analisis sentimen mobil listrik menggunakan metode Naïve Bayes Classifier Sentiment analysis of electric cars using the Naïve Bayes Classifier method. 5, pp.328–335. <https://doi.org/10.37373/infotech.v5i2.146>.
- Burnama, Z.Y., Rosid, M.A. And Azizah, N.L., N.D. Analisis Sentimen Pada Komentar Youtube Dalam Turnamen MPL Season 13 Dengan Metode Ensemble Machine Learning Sentiment Analysis on YouTube Comments in MPL Season 13 Tournament Using Ensemble Machine Learning Method.
- Chely Aulia Misrun, Haerani, E., Fikry, M. And Budianita, E., 2023. Analisis sentimen komentar youtube terhadap Anies Baswedan sebagai bakal calon presiden 2024 menggunakan metode naive bayes classifier. *Jurnal CoSciTech (Computer Science and Information Technology)*, 4(1), pp.207–215. <https://doi.org/10.37859/coscitech.v4i1.4790>.
- Faiq, M., Putro, A. And Setiawan, E.B., 2022. Analisis Sentimen Terhadap Kebijakan Pemerintah dengan Feature Expansion Metode GloVe pada Media sosial Twitter. *e-Proceeding of Engineering*, 9(1), pp.54–66.
- Fitri, E., 2020. Sentiment Analysis of the Ruangguru Application Using Naive Bayes, Random Forest and Support

- Vector Machine Algorithms. *Jurnal Transformatika*, 18(1), p.71.
- Hudha, M., Supriyati, E. And Listyorini, T., 2022. Analisis Sentimen Pengguna Youtube Terhadap Tayangan #Matanajwamenantiterawan Dengan Metode Naïve Bayes Classifier. *JIKO (Jurnal Informatika dan Komputer)*, 5(1), pp.1–6.  
<https://doi.org/10.33387/jiko.v5i1.3376>
- Imran, B., Karim, M.N. And Ningsih, N.I., 2024. Klasifikasi Berita Hoax Terkait Pemilihan Umum Presiden Republik Indonesia Tahun 2024 Menggunakan Naïve Bayes Dan Svm. *Dinamika Rekayasa*, 20(1), pp.1–9.  
<https://doi.org/10.20884/1.dinarek.2024.20.1.27>.
- Informasi, S., Merdeka, U., Jl, M., Raya, T. And Malang, D., 2024a. Analisis Sentimen Pengguna Twitter Terhadap Ai-Generated Art Menggunakan Metode Naïve Bayes Dan Neural Network Kata Kunci : Keyword : pp.51–61.
- Informasi, S.S., Informatika, F., Teknologi, I., Purwokerto, T. and Panjaitan, J.D.I., 2024b. Whitening Dengan Brand Ambassador Exo Menggunakan Metode Naïve Bayes. 03(2), pp.77– 86.
- Jamil, M. And Hanafi, M.S., 2025. Analisis Sentimen Calon Kepala Daerah Maluku Utara dengan Metode CRISP-DM. 7(3).  
<https://doi.org/10.32877/bt.v7i3.2285>.
- Muhayat, T., Fauzi, A. And Indra, D.J., 2023. Analisis Sentimen Terhadap Komentar Video Youtube Menggunakan Support Vector Machines. *Jurnal Ilmiah Komputer*, 19, pp.231–240.
- Pahton, T.Y. And Jati, H., 2024. Analisis Sentimen Data Twitter Terkait Chatgpt Menggunakan Orange Data Mining. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(2), pp.329–336.  
<https://doi.org/10.25126/jtiik.20241127276>.
- Sapanji, R.A.E.V.T., Hamdani, D. And Harahap, P., 2023. Sentiment Analysis of the Top 5 E-commerce Platforms in Indonesia using Text Mining and Natural Language Processing (NLP). *Journal of Applied Informatics and Computing*, 7(2), pp.202–211.  
<https://doi.org/10.30871/jaic.v7i2.6517>
- Saputra, E.F. And Pribadi, M.R., 2023. Analisis Sentimen Komentar Pada Kanal Youtube The Lazy Monday Menggunakan Algoritma Naïve Bayes. *MDP Student Conference*, 2(1), pp.17–23.  
<https://doi.org/10.35957/mdp-sc.v2i1.4283>.
- Sultan, M.I. And Akbar, M., 2025. Analisis Sentimen Pemecatan Jokowi Pada Komentar Publik YouTube Tempo . co. 07(02), pp.125–140.
- Syafia, A.N., Hidayattullah, M.F. And Suteddy, W., 2023. Studi Komparasi Algoritma SVM Dan Random Forest Pada Analisis Sentimen Komentar Youtube BTS. *Jurnal Informatika: Jurnal Pengembangan IT*, 8(3), pp.207–212.  
<https://doi.org/10.30591/jpit.v8i3.5064>
- Winarno, Kurniabudi And Dr. Hendri, S.Kom., S.H., M.S.I., M.H, 2023. Analisis Sentimen Pada Twitter Mengenai Pemerintahan Jokowi Terkait Kesejahteraan Masyarakat Menggunakan Lexicon Based. *Jurnal Informatika Dan Rekayasa Komputer(JAKAKOM)*, 3(2), pp.531–541.  
<https://doi.org/10.33998/jakakom.2023.3.2.846>.